People's Democratic Republic of Algeria
Ministry of Higher Education
University of Abou Bakr Belkaid– Tlemcen

Faculty of Letters, Languages and Arts
Department of Foreign Languages
Section of English

Validating Holistic Scoring for the Assessment of EFL Learners' Writing Performance: Case of Fourth Year Graduate Students at Abou Bakr Belkaid University – Tlemcen

Thesis submitted to the department of foreign languages
in candidacy for the degree of Doctorate in Applied Linguistics

Presented by: Supervised by:

Miss. ZIDANE Rahmouna   **Prof. BENMOUSSAT Smail**

The Members of the Jury:

- Prof. Miliani Mohamed : Prof. (University Of Oran) - President
- Prof. Benmoussat Smail : Prof. (University Of Tlemcen) - Supervisor
- Dr. Bejaoui Fewzia : MC(A). (University of Sidi Bel Abbes) - External Examiner
- Dr. Mellouk Mohamed : MC(A). (University of Sidi Bel Abbes) - External Examiner
- Dr. Hamzaoui Hafida : MC(A). (University of Tlemcen) - Internal Examiner
- Dr. Dendane Zoubir : MC(A). (University of Tlemcen) - Internal Examiner

Academic year: 2009-2010

People's Democratic Republic of Algeria
Ministry of Higher Education
University of Abou Bakr Belkaid– Tlemcen

Faculty of Letters, Languages and Arts
Department of Foreign Languages
Section of English

Validating Holistic Scoring for the Assessment of EFL Learners' Writing Performance:
Case of Fourth Year Graduate Students at Abou Bakr Belkaid University – Tlemcen

Thesis submitted to the department of foreign languages
in candidacy for the degree of Doctorate in Applied Linguistics

Presented by: **Supervised by:**

Miss. ZIDANE Rahmouna   **Prof. BENMOUSSAT Smail**

The Members of the Jury:

- Prof. Miliani Mohamed   : Prof. (University Of Oran) - President
- Prof. Benmoussat Smail : Prof. (University Of Tlemcen) - Supervisor
- Dr. Bejaoui Fewzia   : MC(A). (University of Sidi Bel Abbes) - External Examiner
- Dr. Mellouk Mohamed   : MC(A). (University of Sidi Bel Abbes) - External Examiner
- Dr. Hamzaoui Hafida   : MC(A). (University of Tlemcen) - Internal Examiner
- Dr. Dendane Zoubir   : MC(A). (University of Tlemcen) - Internal Examiner

Academic year: 2009-2010

Acknowledgements

Abstract

Teaching English at the Algerian universities aims at making the students communicate both orally and in a written form. Nevertheless, it can be stated that the writing skill is given more importance since the majority of the examinations revolve around the learners' ability to generate written productions that are often formulated in the form of essays. The assessment of the students' writing performance may incorporate different methods. However, holistic rating is utilized to score the learners' essays.

This research work tries to provide information about the method applied for the assessment of the fourth year students at the English section at Abou Bakr Belkaid University. It makes an investigation about the use of holistic scoring for rating the learners' prompts as well as the difficulties encountered by the students and the teachers in the testing process which is characterized by the employment of essay examinations.

This project relies on the employment of various research instruments including interviews, questionnaires and educational documents that constitute a variety of sources for collecting data.
The analysis and interpretation of the gathered information may be considered as a form of elicitation of the research problematics and the real state of the application of holistic methods for assessing writing.

The purpose of this research work is to diagnose the strengths and weaknesses of holistically scored examinations via the study of the different aspects of writing assessment at higher education as well as the scoring procedure and its impact on the learners' results and the teachers' way of perceiving students' writing. Generally speaking, this project attempts to reveal the extent to which holistic scoring is valid and whether certain kinds of remedies or alterations are needed as far as the testing process is concerned.

CONTENTS

CONTENTS

CHAPTER ONE: **A Literature Review about Assessment**

CHAPTER THREE : CASE STUDY: DATA COLLECTION

CHAPTER FOUR: DATA ANALYSIS AND INTERPRETATION

CHAPTER FIVE: SUGGESTIONS AND RECOMMENDATIONS

APPENDICES:

Appendix A : The Questionnaires

Appendix B : The Questions of the Interviews

Appendix C : The Tests' Questions

Appendix D : A sample of the essays produced by the students responding
to the tests

Appendix E : A sample of students' examination papers

List of Tables

List of Pie Charts

List of Bar Graphs

List of Figures

SYMBOLS

* Erroneous sentences

GENERAL INTRODUCTION

The teaching of English as a foreign language at the Algerian university aims at making the students master this language to get a degree to become teachers of English or use this language for other occupational purposes. The extent of achievement of these aims is checked through the testing of the learners.

The examinations designed for the first, second and third year are meant to assess the students' abilities so that to decide if they will pass to the next level or not. The examinations of the fourth year take place to determine whether the students get a degree or not.

The assessment of the first year is generally based on analytic rating. The majority of the modules of the second and third year are assessed in a holistic way.

The assessment of the fourth year students is totally based on holistic rating because all the examinations involve essay questions. This is why this last group of students is taken as a sample because the main concern of this research work is the study of the assessment based on holistic rating which refers to essay examinations. This work aims at answering the following questions:

1- What are the main issues linked to holistic scoring?
2- What are the difficulties faced by the teachers when correcting essay examinations?
3- What are the difficulties faced by the students when having to answer essay examinations?
4- To what extent are the learners successful in writing essays?

These questions are raised in this research work so that to try to give a view about the difficulties related to the use of holistic scoring for the assessment of the learners' writing performance. Several hypotheses can be put concerning the state of holistic rating. They are as follows:

1- Holistic scoring may raise questions concerning the extent of its reliability and validity as well as the assessors' subjectivity which interferes in such a procedure.

2- The teachers are finding difficulties in measuring students' knowledge and mastery of language due to the non–existence of a clear rating scale which increases teachers' subjectivity.

3- There is a probability that students are encountering difficulties because of certain deficiencies related to the mastery of the language. These difficulties may also arise from the way in which the questions are put forward.

4- The students' writing performance may vary from a learner to another because of some external and internal variables.

The present project work tries to find out to which extent these hypotheses are true. It aims at answering the questions set down as well as providing more information about the process of assessment at the University of Tlemcen focusing mainly on the case of the fourth year students.

This research work is composed of five chapters that are intertwined to enlighten the reader about the research domain. The first chapter provides a theoretical overview of assessment. Thus, some definitions of the concepts of testing, assessment and evaluation in addition to the purposes of assessment are discussed. Next, the chapter deals with the various types of tests, tests' characteristics and effects as well as the concept of accountability. It explains the different steps of test construction and item analysis too. It also describes the main elements involved in the process of testing the spoken and the written language. Then, the chapter moves from the elicitation of assessment issues in general to the description of the concepts related to writing assessment. Therefore, it makes a brief account of the history of writing assessment and tackles the types of tasks as well as the methods used for assessing writing. The different types of scoring and rating scales are listed. Finally, the chapter speaks about the main problems of assessment.

The second chapter focuses on the description of the teaching and testing of English at higher level with special reference to the University of Tlemcen. First, it gives an idea about the concept of teaching English as a foreign language as well as the teaching of the writing skill to EFL learners. It deals with the concept of testing EFL learners' writing performance too. Then, this chapter discusses the situation of teaching English at the Algerian universities. Next, it provides the reader with information about the different modules taught at the section of English at the University of Tlemcen, the general syllabus of teaching this language at this section in addition to the detailed programme of English concerning each graduation level. It also speaks about the methods used for the evaluation of the students. Finally, it describes the process of testing English at university including the features of oral and written examinations as well as the different types of questions utilized in the examinations of each module.

The third chapter concerns the data collection. First, it explains the objectives of the case study. Next, it describes the research design by eliciting the purpose of the approach employed, the different phases of the research work and the techniques used for the analysis and interpretation of the gathered information. Then, the chapter gives a description of the sample population and the methods of sampling as well as the variables involved in the case study. It also lists the research instruments utilized for collecting data and states the goals and advantages of employing such tools. Theses instruments include interviews, questionnaires which are designed to be answered by the students and the teachers, tests and instructional documents. The aim of this chapter is to provide a general overview to the readers about the overall framework of the case study by revealing the nature of the research methodology that governs the process of data collection which represents the core of the present research work as it establishes a basis for the kind of results and conclusions drawn from the investigation taking place concerning the use of holistic scoring for writing assessment at the Algerian university.

The fourth chapter deals with the analysis and interpretation of the collected data. Thus, the information provided by the interviews and the questionnaires are analysed to present a global view about the teaching and testing situation. The results of the utilized tests are listed for the purpose of understanding how holistic scoring is applied for assessing the essays produced by fourth year learners. Next, students' examination papers are studied to elicit the type of mistakes produced by the learners and the students' way of writing compositions. Then, students marks are analysed in order to describe the students' degree of achievement. After the analysis of the data obtained from the various instruments, the results are interpreted via descriptive statistics and qualitative evaluation of the gathered information. Finally, the main research results are discussed by making reference to the findings of researchers and explaining the degree of evidence and generalizability of the collected data.

The fifth chapter provides some suggestions and recommendations related to the teaching and assessment of the writing skill. Such suggestions focus on increasing the efficiency of essay examinations, maintaining and improving holistic scoring in addition to the utilization of analytic marking scales for rating compositions and the construction of scoring rubrics. This chapter also discusses the importance of employing a set of strategies for assessing dissertations as well as providing feedback via assessment. Distributive evaluation, product analysis and portfolio assessment are explained for the sake of recommending their use for developing the testing process. Then, the chapter lists the advantages of exploiting the six traits of writing for teaching and assessing this skill. Finally, some suggestions concerning the instruction of writing are put forward such as the use of assessment as a tool for teaching this skill and raising the learners' motivation as well as the employment of a variety of techniques to teach writing.

In fact, this project aims at describing the situation of writing assessment through the utilization of various instruments that give the possibility to get a wide variety of information about the subject of investigation. However, the studied

population will be represented by a set of samples whose size is limited due to the difficulty of time management and data analysis. Therefore, the collected data will be considered as a reflection on the whole state of the target population.

The present research work focuses on collecting information about how teachers deal with essay examinations and the difficulties they find in correcting students' compositions. It attempts to investigate about scoring consistency by trying to examine how the same paper is scored by different teachers or by the same teacher in different occasions. It also studies the main mistakes produced by the learners and the way the scorers refer to these mistakes. In general, the case study puts the emphasis on the use of holistic scoring for assessing writing taking into consideration the different aspects of this procedure and its effect on the learners' performance.

The analysis of the collected data will rely on employing both qualitative and quantitative information for the purpose of interpreting the obtained results. The interpretation of the gathered data will attempt to provide a sort of evidence to the research hypotheses and answers to the research questions. The conclusions drawn from the research results will be taken as a basis for the provision of a set of suggestions that may be beneficial for the teachers.

Since teaching and testing are interrelated, the recommendations that will be provided in this project will concern the different strategies that may improve the process of teaching writing as well as the process of writing assessment. This research work gives importance to the procedure of holistic scoring as it is considered as the main direct method of writing assessment. In addition to this, the majority of the teachers employ such a technique for rating their students' essays. Consequently, this testing instrument has a strong impact on the quality of the judgement made about the learners' performance. Therefore, it is essential to shed light on holistic scoring for the sake of evaluating the conditions of the assessment of writing at university.

CHAPTER ONE:

A Literature Review
of Assessment

CHAPTER ONE:

A Literature Review about Assessment

1.1- Introduction

Teaching and testing are two closely related concepts as the latter provides information about the effectiveness of the former. Testing is also referred to as assessment. This work aims at discussing the concept of holistic scoring and the difficulties related to it; but before doing so, there must be a sort of literature review which summarizes all the important issues related to the field of testing and assessment.

The first chapter tackles some of the theoretical studies concerning the purposes and means of assessing the learners' performances. It will provide definitions concerning testing and assessment in addition to the extent of similarity and difference between them. It will also speak about the different types and purposes of assessment as well as the different methods used for testing and the qualities of a good assessment.

1.2- The Definition of Testing, Assessment and Evaluation

The terms testing and assessment are most of the time used to refer to the same concept. In most cases, they have the same meaning. But sometimes, the term assessment may have other meanings than testing. In this research work, the terms testing and assessment will be used interchangeably.

1.2.1- Testing

Testing refers to the process of measuring the learners' abilities and performances through the use of examinations and tests in order to *"...provide information for making a variety of decisions"* (Bachman, 1990: 70). It aims at analyzing the learners' levels and their degree of success in learning. In this context, the tests' results are used to differentiate between the performances of the learners and to make a decision concerning whether they will be allowed to pass from a level

to another, get a degree, follow a certain course of study, be placed in a certain class or be considered as unsuccessful.

Testing can take place during the teaching programme. In this case, it is called formative testing. When it takes place at the end of the teaching programme, it is called summative testing.

> *Summative testing comes at the end of a course or program of instruction and helps teachers make decisions about passing, failing or promotion or about the overall effectiveness of the program or course"* (Genesee and Upshur,1996 : 153)

Thus, summative testing is concerned with the degree of achievement of the educational objectives. It takes place at the end of the instructional programmes to make an evaluation of the learners' performances as well as the effectiveness of the teaching programme. On the other hand, formative testing focuses on the kind of achievement or weaknesses that occur during the teaching process. *"Formative assessment is assessment that is meant to guide both teachers and students toward the next steps in the learning process"* (Benjamin, 2008:3).

Formative testing represents a sort of continuous assessment since it occurs within the teaching courses. It aims at giving information about the kind of progress achieved in making the learners master the elements that were taught to them as well as the difficulties they find when learning. These data will enable the teacher to proceed with the following courses or to make a revision before starting a new course. Formative testing is also referred to as assessment for learning while summative testing is called assessment of learning (Stiggins et al., 2004).

Generally speaking, summative and formative testing refer to the classification of assessment into types according to the period it takes place in. On the other hand, testing can be classified according to the way used by the scorer in order to assign a mark to the learners' answers. This is why testing can be objective or subjective.

> *If the examiner has to exercise judgement; if he has to decide whether the answer is adequate or inadequate; if he has to choose between awarding it a high or low mark; then the marking process is subjective. If on the other hand, he is precluded from making judgement... if in short, he is reduced for the purpose of marking, to the status of a machine; then the marking process is objective.*
>
> (Pilliner, 1968: 21)

Testing is objective if the scorer does not give his opinion and awards marks according to an established marking scale. If the scoring is based on the scorer's judgement, the testing process is subjective.

Subjective testing requires the scorer's personal opinion concerning the test answer. It relies on the degree of his appreciation of the learner's performance. In this sense, subjectivity is linked to the interference of the scorer's judgement in the scoring process.

Objective testing refers to the type of assessment that is based on a definite marking scale which accepts only one answer for each test question. The objectivity of the testing process requires an automatic scoring from the part of the scorer without an interference of his own personal judgement since all the answers are considered either as true or false and are given precise marks.

Assessment can be subjective or objective according to the way used by the rater to rate the test. *"The distinction between a subjective and objective examination rests only on the manner in which the marks are to be assigned"* (Pilliner, 1968: 21). It is the method used for scoring that determines if testing is objective or subjective.

Assessment takes different forms according to the function it has to fulfil. Other types of testing exist including internal, external, outcomes and programmatic testing. Internal assessment focuses on the individual's learning and is tied to the curriculum. External assessment is not fixed to a particular class and is used to compare student achievement among districts. Outcomes assessment puts the

emphasis on process learning in depth and problem solving within a context. It aims at measuring students' achievement of the goals designed for a specific course. Programmatic assessment provides information about the success of educational program and how well it is meeting its goals (Lippman, 2003).

Assessment has many audiences called stakeholders that refer to the people who hold some stake in the decisions of testing. These audiences are classified into three levels. The primary level of stakeholders is the learners, their teachers and their families. This category includes the highest stakes as they are concerned with the testing situation and results. The secondary stakeholders refer to those persons who are not directly involved in the assessment process but they are in touch with this context like administrators. The tertiary level of stakeholders means the people who have little or no contact with the testing situation but they may interact with the results of the test (Lynch, 2003).

The aim of testing is to get sufficient data concerning the learners' performances. The type of information provided from the testing process will enable the tester to draw certain conclusions about the learners' abilities and the kind of decisions that should be made.

1.2.2- Assessment

Most of the time, assessment refers to testing because it provides information about the learners' abilities for the sake of making certain decisions. Assessment refers to the process of gathering information about students from various sources to enable the teacher to understand these learners and describe them. Therefore, assessment gives a description of the students' performances in order to provide the possibility to make a sort of judgement about these learners and the kind of decision making which will lead to classify them into groups, to allow them to study at a certain level or to give them a certificate or degree.

Assessment can provide information about the learners through the use of various means. It takes place when a test is administered to learners since it *"...is a measurement instrument designed to elicit a specific sample of an individual's*

behavior" (Bachman, 1990: 20). This means that a test is an assessment tool that enables to get data about the learner's performance and his achievement as far as the learning situation is concerned. In this way, assessment is synonymous to testing as both of them are considered as sources of information concerning learners which will be used to make decisions that will affect the learners and the learning process.

> *Although testing and assessment are often used interchangeably, the latter is an umbrella term encompassing measurement instruments administered on a 'one-off' basis such as tests, as well as methods of monitoring and recording student learning such as observation, simulations, or project work.* (Brindley, 2001: 137)

Sometimes, assessment can have a different meaning from testing. There can be an assessment of the learners' abilities without a test when the teacher observes the behaviour of the learners in the classroom from their way of participating and doing the exercises. Checklists may be used to make each learner answer a set of questions to get information about his learning difficulties.

Even if assessment can take place without a test, it is very often preferable to use the test as a source of assessment because it provides exact information through the test's results which will enable to make decisions from the obtained marks.

Therefore, most of the time, the term assessment refers to testing. Assessment or testing is defined as the measurement of the learners' abilities and performances that represents a sort of data gathering for the sake of decisions' making which is a part of the process of evaluation.

1.2.3- Evaluation

Evaluation is the process of drawing conclusions about the learners and the learning situation from the data collected about them. *"Evaluation can be defined as the systematic gathering of information for the purpose of making decisions"* (Bachman, 1990: 22).

It can have different sources since this type of information can be collected by several means. When information about the learners is elicited through tests, the evaluation process is taking place in the form of assessment or testing which focuses on the study of the learners' performances to come out with decisions concerning them. Thus, there is a sort of investigation concerning the learners' abilities.

On the other hand, evaluation can have a wider sense than assessment and testing when it is concerned with gathering information about the syllabuses, the textbooks, the teaching materials and the teaching methods. In this case, tools such as observation and questionnaires are used instead of testing. *"Testing is only part of evaluation procedure. We need measurements of all the variables: teacher, learner and materials"* (Corder, 1973: 352).

Suskie and Banta (2009) suggest three main definitions for evaluation. The first perspective links evaluation to the interpretation of the results of assessment through the use of the information provided to make a judgement about the learner' achievement of the established learning goals as well as the strengths and weaknesses of teaching and learning strategies in order to suggest appropriate changes in the goals and teaching/learning strategies.

The second definition implies that evaluation makes a link between the actual outcomes and the intended outcomes of the assessment process. In this case, the assessment and evaluation of student learning are seen as synonymous.

The third conception of evaluation refers to the investigation into the quality or worth of a program rather than student learning. This definition means that evaluation is a broader concept than assessment.

Therefore, evaluation refers to the global measurements of all the aspects involved in the learning and teaching process through the use of tests and other means of investigations to get information not only about the learners but also about the teachers and the teaching materials. In fact, evaluation, testing and assessment may differ in their meaning according to the purposes they achieve.

1.3- The Purposes of Assessment

Assessment can have several purposes such as selection, diagnosis and the evaluation of the teaching process.

1.3.1- Selection

Assessment helps to determine the learners' level and to decide if they pass or fail. Thus, its main purpose is the selection of individuals for educational opportunities or careers (Rowntree, 1987).

Sometimes, assessment is used to assign the testees to specific sections according to their level of competence (Harris, 1969). In this case, the aim of testing is to place the learners in certain classes on the basis of their actual abilities or to allow them to follow a specific course of study. *"Assessment is obviously an important part of any course since it in fact determines what the student must do in order to gain the qualifications"* (Wallace,1991 : 126).

This means that assessment represents a source of information about the students' abilities which will enable the testers to make decisions concerning whether they are qualified to continue a certain course or not.

1.3.2- Diagnosis

Learners can be assessed in order to get informed about their degree of achievement in learning. The testing process takes the form of a diagnosis of the learners' abilities. Tests help *"... to assess students progress and to assist in evaluating students achievement"* (Bachman, 1990: 3). Assessment enables the testers to be aware about the occurrence of a sort of improvement in the level of the learners or to measure the students' knowledge of a certain skill or some elements of a language. Thus, it leads to get information about the students' achievement concerning the learning process either by eliciting their performance in general or by providing a sample of their performance in a certain domain of knowledge.

1.3.3- The Evaluation of the Teaching Process

Assessment helps to evaluate the teaching process since it provides data concerning the kind of progress made in achieving the objectives designed for the teaching programme. *"Language tests can be valuable sources of information about the effectiveness of learning and teaching"* (Bachman, 1990: 3). Thus, assessment gives an analysis of the teaching situation so that the teacher will know if the learners have mastered what was taught to them or not. It may indicate learners' achievement and the difficulties they encounter when learning as well as the problem areas embedded in the language syllabus (Heaton, 1975).

The testing process can lead to understand the learners' weaknesses. It enables the tester to know the kind of deficiencies linked to the teaching situation and the problems encountered by the learners in their learning process. This diagnosis of the learners' weaknesses will help to make a remedial work and to bring new techniques to improve the teaching situation. Since assessment can provide different types of information concerning the learners, it involves the classification of tests into various categories.

1.4- Types of Tests

There are different types of tests according to the tests' content , the methods used for testing , the kind of measurement used , the type of information provided by the tests, the kind of decisions made from their results and according to the way they are constructed.

1.4.1- Types of Tests according to Content

The content of a test involves a sample of the subject matter or the knowledge which is going to be tested. Tests can be classified according to their content into three types which include discrete-point tests, integrative and communicative tests.

Discrete–point tests are designed to test each skill on its own or to test certain elements of the language such as grammar, vocabulary, spelling or pronunciation.

> *Discrete-point testing concentrates on testing separately different language items and language skills (eg. Grammar, sound discrimination , listening with comprehension , writing) and by combining the results of a number of separate test items builds up a picture of the student's level of English.*
>
> (Cunningsworth, 1984: 56)

This means that discrete-point tests focus on a certain skill or a specific aspect of language without taking into consideration the remaining skills or elements, but in reality, "*...tests are seldom truly discrete points in nature because no test or test item depends solely on one subskill to the exclusion of all others*" (Genesee and Upshur, 1996: 152). As this kind of tests involves only one skill, it is rarely used. Most of the time, the content of the test involves an integration of the skills or some of them.

Integrative tests focus on testing the skills in an integrative way. These types of tests involve the assessment of the testee's use of several elements of a language at the same time.

> *Tests that call on a number of subskills operating in concert are sometimes referred to as integrative tests ...Performance on a grammar test for example can also reflect the test taker's spelling, vocabulary or even organizational skills with language.* (Genesee and Upshur, 1996:152)

The content of integrative tests involves a variety of questions related to different aspects of the language that aim at testing the learners' mastery of various skills. The most common activities included in such tests are dictation and the cloze test (Lynch, 2003).

Communicative tests include subjects linked to real situations requiring the learners to use different skills and language forms. These test types require the students "...*to perform in a communicative situation using whatever combination of skills and abilities is necessary*" (Cunningsworth, 1984: 56). Therefore, communicative tests focus on making the learners recall all the aspects of a language and try to express themselves meaningfully and correctly in a situation which resembles authentic life. Their content tries to check if the learners are able to communicate using the target language.

1.4.2- Types of Tests According to the Kind of Measurement

These tests refer to the way of measuring the learner's performance. They are: norm-referenced tests, criterion-referenced tests, domain-referenced test and objectives-referenced tests.

In norm-referenced tests, the learners' performances are measured by making a comparison between the kind of performance of a learner or a group of learners and another group of learners. "*A student's score is therefore interpreted with reference to the scores of other students or group of students rather than to an agreed criterion score*" (Richards, 1990: 16). These tests measure the extent of the learners' ability to answer the test questions by drawing similarities and differences between a student's grade and those grades of other learners so that "...*the total number of candidates is assumed to fall into some kind of normal distributions, in which there will be a few very able people, a few very weak people, and the majority scattered at various points in the middle*" (Wallace, 1991: 132). Therefore, the kind of expectation made about the different performances of the learners will be taken as a norm in order to find out if their results reflect their individual differences.

Criterion–referenced tests set a criterion as a basis for the measurement of the performance of the learners. To pass the test, the learner must reach this level of performance that is considered as a standard (Richards, 1990).

These tests do not involve any kind of comparison between the scores of the learners. *"The measurement of pass or fail is not done against some internalised judgement on the part of highly skilled examiners, but by some observable and measurable criteria which can be demonstrated in the candidates' performance"* (Wallace 1991: 132). Therefore, the criterion designed for measuring the learner's performances may involve his ability to master a certain specific skill or to use the language to elicit his understanding of a certain course or domain of study. The learner will be considered as successful or not according to the degree of his readiness to perform well in the test and reach the level which is determined by the established criterion.

Domain–referenced tests measure the learners' performances with reference to a domain of knowledge. They are related to how learners perform when required to show their ability as far as using the language in a specific situation by assessing the learners' mastery of a certain skill or field of study. *"Domain-referenced tests provide interpretations of test scores relative to an identified domain of knowledge or skill"* (Genesee and Upshur, 1996: 150). For instance, this domain of skill or knowledge may refer to the learners' ability to read a certain passage, to use the language in a written form or to express one's self fluently within a certain context.

Objectives-referenced tests focus on the degree of achievement of the educational objectives. They involve the measurement of the performance of the learners and the interpretation of the test scores with reference to the instructional goals designed to be attained. These types of tests aim at getting information about the extent of the learners' mastery of what has been taught to them and whether the objectives set for instruction have been reached or not.

1.4.3- Types of Tests According to the Information Provided

Some tests are designed to provide information about the learner and the learning situation. This type includes achievement tests, proficiency tests, performance tests, progress tests, diagnostic tests and aptitude tests.

Achievement tests, also called attainment tests, are designed to assess the extent of the learners' mastery of what has been taught to them. The purpose of such tests is to know if the learners master the elements studied since they give data about student attainment. *"An achievement test is in effect a sample of all the kinds of tasks that a given course of study is striving to get students to master"* (Dyer, 1964:456).

These tests enable the tester to have sufficient information about the kind of achievement made by each learner and the degree of improvement in the learning process.

Proficiency tests are employed to measure the learners' capacity to utilize the language in a specific context in order to determine their readiness to perform well in a future learning situation. Thus, they help the tester to decide if the learners have the capacity to express themselves using the target language. *"The aim of a proficiency test is to assess the student's ability to apply in actual situations what he has learnt"* (Harrison, 1983: 7). These tests are designed to know the kind of learners' future performance and their ability to put what they learnt into practice.

Performance tests aim at providing data about how well the learners are going to perform and if they are able to express themselves in a present situation which is similar to real life. They elicit information about students' ability to employ the language to perform authentic tasks (Genesee and Upshur, 1996).

They measure the quality of the learners' performance by investigating the way they behave and how they use the taught language to cope with a certain authentic situation.

a decision about their ability to perform well when following a certain type of language instruction.

1.4.4- Types of Tests According to the Undertaken Decisions

Some tests enable to make some decisions from their results. These tests are placement tests, screening tests and final examinations.

Placement Tests focus on testing the learners' general ability for the purpose of placing them into groups of the same level to start learning a certain programme. They are designed *"...to assess an applicant's language level so that he can be placed in a class suitable for his abilities"* (Carroll, 1980: 75). These tests are used to provide information about the students' degree of mastery of the language to determine the appropriate type of instruction for these learners. Their purpose is to classify the learners into groups according to their levels.

Screening tests aim at enabling the learners to follow certain studies if their level is good or to refuse to let them follow such learning if their performance does not correspond to the required level. They give information about the learners' abilities so that the tester can make a decision concerning whether these learners follow a certain course or not. Screening tests are utilized to distinguish those who are prepared for an academic program from those who are not (Harris, 1969). They make a kind of distinction between the levels of the testees in order to enable them to continue studying or change the domain of learning.

Final examinations are those types of tests which lead to decide if the learners succeeded in learning an educational program or they failed in reaching the required level for being considered as successful. *"Final examinations are used to promote students to the next level of instruction or to award a certificate of success or completion"* (Genesee and Upshur, 1996: 153). Therefore, they aim at determining the learners' success or failure in their learning process. From their

results, a decision is taken in order to lead the students to be allowed to continue studying in the following level or get a degree.

1.4.5- Types of Tests According to their Construction

Tests can be constructed by teachers or by certain institutions. This is why two kinds of tests are distinguished according to the manner of their construction. These types of tests are teacher-made tests and standardized tests.

Teacher made tests, also called classroom tests or non-standardized tests, are constructed by the teacher who is in charge of teaching a certain class. Thus, the same individual plays the role of an instructor, test writer and evaluator (Harris, 1969). These tests focus on assessing the learners' mastery of what was taught to them during certain courses.

Classroom tests are made and scored by the same teacher who aims at diagnosing the learners' performances and the degree of his achievement of the educational goals. *"Teacher-made tests are devised to assess achievement as a result of exposure to specific instructional activities"* (Genesee and Upshur, 1996: 235). They are based on the content of certain specific courses which leads the learners to be aware about what is asked from them and be at ease to a certain extent when answering the test because they are more accustomed to their teacher's testing methods.

These tests are constructed, administered and rated by the same teacher who is conscious about his learners' abilities and level of performance. They aim at diagnosing the learners' achievement and degree of progress in learning.

Standardized tests are constructed by certain institutions or a group of teachers. They are used to test the learners for the purpose of allowing the successful ones to be placed in certain classes or to get a degree. Standardized tests *"... are prepared by professional testing services to assist institutions in the selection, placement and evaluation of students"* (Harris, 1969: 1). They are designed by specialized people who are in charge of writing tests that aim at assessing the

learners' general performance so that to make a judgement about their success or failure. Unlike teacher-made tests which aim at checking the learners' mastery of specific courses, standardized tests try to provide information concerning the learners' knowledge and their ability to use the target language to communicate either orally or in a written form.

The testing process involves the use of different types of test that measure various aspects of the learners' knowledge which means that the purposes of assessment vary from one situation to another. However, all the types of tests must possess a set of characteristics that determine their usefulness.

1.5- Tests' Characteristics

Tests are the means of measuring the learners' performances. They include a set of questions or exercises which have to be answered by the students in order to give them certain grades. They are a form of measurement. They attempt to give information about the learners' degree of achievement in the process of education. Examinations are also referred to as tests because they have the same purposes of the latter. *"Tests have subject matter or content . . . a test is a task or set of tasks that elicits observable behaviour from the test taker ... Tests yield_scores that represent attributes or characteristics of individuals"* (Genesee and Upshur, 1996:141).

Thus, tests should have a certain content and involve certain tasks or exercises that should be undertaken by the learners. They provide results or information concerning the learners' performance in the form of scores. Tests must have some characteristics that make them really effective. These features are validity, reliability, discrimination and practicality.

1.5.1- Validity

Validity refers to the fact that the test has to measure the elements it has been designed for assessing. It is related to the extent to which the test succeeds in measuring what it was administered for. *"The validity of any examination or test procedure may be broadly defined as the extent to which it does what it is intended to do"* (Pilliner, 1968: 30). It is linked to the content of the test and its suitability to the testing situation.

The test validity is closely related to the idea that what is tested is reflected in the test results. This concept concerns the truth-value of a test and its scores (Davies and Elder, 2005). In order to be valid, the test must show the degree of resemblance between the performance of the learners and the objectives set to be achieved from the content of the test. Cohen et al (2007) mention different types of validity such as face validity, content validity, empirical validity, criterion-related validity, construct validity and consequential validity.

Face validity is linked to the appearance of the test and people's opinion about it. In order to possess face validity, the test should be relevant and appropriate to the testing situation. It has to be considered as acceptable by the learners and it should not decrease their motivation. Face validity refers to the extent of the test relevance and its estimation by those who are concerned by the test whether testees or testers.

Content validity concentrates on the measurement of the content of a particular course of study and the aim wanted to be achieved from it. This is why there must be a certain choice of the test items in order to get information about the specific materials or skills being tested (Harris, 1969). Content validity measures the extent to which the test content really tests specific samples of the teaching programme.

To measure content validity, the elements meant to be tested have to be clearly determined so that to get exact information about the degree of the learners' mastery of that specific content.

The notion of face validity and content validity lead the concept of validity to be synonymous with the authenticity which refers to the extent to which the test tasks reflect the language used in actual situations. *"Face validity is the appearance of real life; content relevance is the representation of real life"* (Bachman, 1990: 307).

Both face and content validity focus on connecting the test's tasks to reality. Thus, they refer to authenticity. Although both types of validity are linked to authenticity, the latter seems more closely related to content validity as it deals with the relevance of the tasks of the test to the target language use domain (Bachman and Palmer, 1996).

Empirical validity refers to the degree of similarity between the test results and a certain criterion which is set as basis for comparison. There are two types of empirical validity. The first kind is called predictive validity and the second type is called concurrent or status validity.

Predictive validity measures the extent of correlation between the test scores and a subsequent criterion which means that the marks of the learners in a certain test are compared with the marks obtained in a test of a previous term or semester. Concurrent validity refers to the measurement of the degree of correlation between the test scores and a concurrent criterion (Williams and Burden, 1997). In this case, two tests are administered immediately one after the other and a comparison is made between the scores of the first and the second test.

Criterion-related validity refers to the type of correlation which exists between the results of two tests and a certain attribute or quality (Genesee and Upshur, 1996). For instance, if some learners are placed in a certain group or class according to the results they obtained in a specific test, they should perform well

when being in this group. The results they obtain in a test administered in their new situation are compared with those obtained in a test administered before their placement in this group. If the first results correlate with the second ones according to the established criterion, the placement test is said to have criterion-related validity.

Construct validity indicates the degree of agreement between a certain established construct concerning a test and the kind of information provided by this test. A theory is set as a sort of prediction for the test results. The way these results relate or correspond to that theory will give evidence if the test possesses construct validity or not (Williams and Burden, 1997; Genesee and Upshur, 1996).

Construct validity is closely related to the notion of interactiveness which is defined as *"...the extent and type of involvement of the test task"* (Bachman and Palmer, 1996: 25). Thus, the degree of correspondence between construct validity and high interactiveness relies on the characteristics of the testees and the way of defining the construct.

Consequential validity is linked to the impact of a test on the educational and social systems. It refers to the kind of consequences resulting from the use of the scores obtained in a given test (Hyland, 2003).

Although there are various types of validity, one can say that validity is generally related to the content of the test. Davies and Elder (2005) claim that research about validity is manifested through two approaches which are positivism and interpretivism. Positivist validity correlates between the test and an agreed criterion so that to ensure if the relevant trait or construct is being measured. Interpretivist validity investigates about the nature of the element that is tested.

The process of investigation about validity is called validation which refers to the act of making studies in language testing for the sake of bringing proofs for the validity of the tests employed. It involves a statistical analysis of test scores in addition to the use of qualitative and quantitative approaches focusing on the nature of the test and the qualities of the performance elicited from it. The techniques of

validation also include discourse analysis which is employed to study the speech elicited in oral proficiency interviews as well as the test discourse (Davies and Elder, 2005).

Another characteristic of a good test, in addition to validity, is reliability.

1.5.2- Reliability

Reliability refers to the stability of the test results. It means that approximately the same results are obtained by the learners when tested several times and under different circumstances. *"On repeating the procedure, the scores should not be subject to fluctuations, and hence, the persons obtaining these scores should maintain their standings relative to one another"* (Pilliner, 1968: 23). The reliability of a test depends on certain factors like the number of the samples of the students' performance, the conditions under which the test is administered and the learners' motivation (Harris, 1969).

Thus, the test should assess different aspects of the learner's abilities and should take place under similar conditions so that to possess reliability. The learners' motivation is also considered as a very important factor that deeply affects reliability since high and continuous motivation will cause the test to be reliable while low motivation causes a low reliability. But this motivation cannot be controlled since sometimes the questions of the test may be poorly designed which will decrease the learners' motivation or some of the learners may have a low motivation because of some psychological or personal problems.

Various types of reliability exist according to the method used for its measurement. The different types of reliability are: test-retest reliability, parallel or alternate forms reliability, internal consistency reliability and scorer reliability.

Test–retest reliability involves administering a test to the learners twice with a certain interval between the first and the second time (Genesee and Upshur, 1996). If those scores are found to be approximately the same, the test is said to be reliable. In this case, the reliability of the test depends on the time interval existing between the first and the second test. Thus, this period of time interval should not be very short nor very long so that the learners' performances would not largely differ. Test-retest reliability is also called the coefficient of stability.

Parallel or alternate forms reliability refers to the stability of the learners' results when having the same test administered several times with slight changes in its form. The students have to answer the first test which is considered as the basis for measurement. Then, they have to sit for other tests whose versions are different but they are parallel to the first one in length, difficulty and time limits (Harris, 1969). The scores obtained in the first test and the results of the tests that have equivalent forms of it are compared. The high correlation between those scores refers to the reliability of the test. Alternate forms reliability focuses on those errors linked to the timing and the content of the test.

Internal consistency reliability, which is also called split-half reliability, is linked to the stability of the scores concerning the different items of the test which is administered just once but its components are separated into two parts. Each part is scored on its own so that to give each learner two scores which will be compared later on (Genesee and Upshur, 1996). Thus, internal consistency is concerned with the extent of correlation between the scores of the different elements of a test and the consistency of the measurement taking place via these elements or tasks.

Scorer reliability is also called rater reliability. It is related to the stability of the way of scoring the test (Harris, 1969). It refers to the fact that approximately the same scores are assigned by one scorer when the test is repeated, or several scorers give approximately the same scores to the learners in a given test. Scorer reliability is referred to as interrater reliability when the test is scored by several assessors; it is

The standard error of measurement shows to which extent the test is reliable and whether the scores reflect the abilities of learners or not. The test reliability is very essential in determining the stability of the scores. Interpretivist approaches to test validity claim that reliability is not as necessary as validity since the former offers generality that is repetitive and recursive while the latter offers newness and change (Davies and Elder, 2005). However, a good test has to be both valid and reliable. In addition to this, it should possess another characteristic which is discrimination.

1.5.3- Discrimination

Discrimination means that the test has to show the differences which exist between the learners. *"An important feature of a test is its capacity to discriminate among the different candidates and to reflect the differences in the performances of the individuals in the group"* (Heaton, 1975: 165). The test must give a description of the learners' abilities and the way each learner performs differently from the other learners. This is why the learner's scores should provide clear information about his/her performance. Pillliner states that:

> *The purpose of the instrument whether examination or test is to reflect differences in the attribute concerned which are presumed to exist among the subjects... The extent to which a single subject manifests the attribute becomes meaningful only in relation to the corresponding manifestations by others.*   (1968: 23)

Thus, the test should enable the tester to know the degree of the learner's ability to use the language elements such as vocabulary or grammar as well as the degree of the divergence of his performance from the other learners. When compared with each other, the learners' performances should reveal certain differences. In this way, the test will discriminate between the levels of the learners

according to the distinction existing in the scores. However, the test should not only be characterized by discrimination but also by practicality.

1.5.4- Practicality

Practicality, or usability, indicates the degree to which the test can be put into practice through the use of means and provided facilities that are referred to as practical considerations. These considerations involve economy, ease of administering and scoring as well as the ease of interpretation (Harris, 1969).

Economy is related to the cost and timing of the test. Economy in the cost of the test means that the test should not require much money to be administered and has to be economical as far as the cost of the copies and the scoring process are concerned. In addition to economy in money, the test also has to provide economy in time by possessing sufficient length. It should not be too short to enable the tester to have enough information about the learners' abilities; it should not be too long so that not to become a burden for both the learner and the scorer. Therefore, the test has to be set according to a certain time limit which is considered as acceptable. This economy in time will enable the testees to be at ease when answering. Also, the scorers will not find many difficulties in correcting.

Economy in the cost and timing of the test lead to its practicality. The latter is also provided when its administration and scoring are facilitated.

Another feature of practicality is the ease of administering and scoring the test. In order to be practical, the test application must be easily held so that the test administrators and the test takers will not find difficulties when the test is conducted. All the conditions must be favourable. Facilities concerning the test scoring must also be supplied. The way of scoring the test will determine the degree of difficulty of the scorers' task. The test questions should not have vague answers so that the scorers do not face problems in deciding about the mark to be awarded.

In addition to the ease of administration and scoring, the practicality of the test depends on the ease of the interpretation of the test results. Thus, the test scores must give sufficient information about the learners' abilities so that to enable the test administrators to make decisions according to the learners' performances. The test results should be easily interpreted to be used as a basis for the analysis of the learning and teaching situation for the sake of improving the testing and the teaching process. From the scores, the test validity and reliability can be measured in order to give data concerning the learners. This is why the test results should provide a clear description of the learners' performances so that to facilitate their interpretation.

Therefore, all tests must possess validity, reliability, discrimination and practicality in order to be considered as being good ones. Each designed test can differ from the other ones in its content, its construction and its purpose. However, all tests have effects on the learners and the society.

1.6- The Effects of Tests

Tests have a great impact not only on the learners but also on their parents, the teachers, the administration and the whole society. They are used to get information concerning the learners in order to make decisions that deeply affect their lives and cause them to follow a certain domain of study or occupation.

The effect of tests is known as test washback or backwash which refers to the fact that any test is going to affect the learning and teaching situation as well as the future of the learners (McNamara, 2000). The test results will determine if the testees are going to pass or fail. They will be taken as a basis for deciding whether they will be accepted or rejected from a certain course of study. Tests also enable the learners to get a degree or a diploma so that they can get a job. The parents of the learners are also concerned by the effects of tests since they worry about their children and want them to succeed in their studies. The test results will either please or dissatisfy them. Tests have effects on the teachers who will get an idea about the degree of the effectiveness of their teaching from the tests' results. They also affect

administrations and schools which will have a good or bad reputation according the kinds of results obtained.

Tests have an effect on the society as a whole since they determine the future of the testees and the validity of the educational programmes and methods.

> *Washback can be beneficial, as in the case where changing or instituting language measures leads to beneficial changes in teaching and curriculum. Washback can also refer to negative effects, such as when the testing program fails to recognize course goals and learning objectives to which the test is supposed to relate.* (Salaberry and Cohen, 2006:158)

The effect of the test may be positive or negative. Positive or beneficial washback is linked to the influence of tests' results on the curriculum which may be altered to suit the learners' needs. Negative washback means that the test allows the students to pass without preparing them for the demands of the target language use tasks (Moore and Morton, 2007).

There is a distinction between the terms test impact and washback. The test impact includes washback as it refers to effect of tests on the community. It is synonymous with consequential validity. However, washback is linked to the influences of tests on teachers, teaching and learning (Shaw and Weir, 2007).

Thus, the tests can have a great impact on the learners and the process of instruction. The test effects are related to educational accountability.

1.7- Accountability

Accountability means that the testers should take into account the idea that the test's results have to be satisfactory to a certain level. They must think about the kind of effects the test will have on the learners, their parents and the educational environment. School managers, teachers and educators are responsible for the results obtained from tests.

Accountability is a term that has been employed since the 1970's. _"This concept comes from an accountant's ledger that all too often places the budget at the center of the decision-process"_ (Rosano et al., 2005: 121). Therefore, accountability relates the budgets to the instructional needs to make positive changes and achieve educational success. However, it often links the results of the tests to the teaching goals, the instructional programmes and the teaching/learning outcomes. First, accountability can be used to refer to the degree of success of the educational institution to realize the established teaching and testing purposes. Second, it may be associated with the efficiency of the educational curricula and determine whether to keep or change them on the basis of the existing results. Accountability can refer to the consistency of the assessment outcomes with the educational objectives and programmes. Therefore, the tests provide information about the success or failure of the learners and the institution. In addition to this, they determine the future of the students and compel the teachers to relate the teaching process to the assessment purposes.

Accountability in education is given a great importance since the cost of education which is part of the expenditure of the country is considered as an investment that aims at forming well educated students who will contribute in the development of the society. This means that assessment and accountability are interrelated as the former is a tool of measuring the success of education which is closely linked to the future of the learners and the educational system (Harris, 1993). Accountability is often connected to instructional achievement. It is generally defined as the responsibility of the educational system for the results of the tests.

> _Responsibility, while similar to accountability, can be perceived of as more inclusive and places the answerability for the success or failure of young people's learning on all of society- the public, legislators, parents, teachers, administrators as well as the schools._
>
> (Rosano et al., 2005: 121-122)

This concept urges all the people of the society to collaborate in order to improve the educational process and enable it to achieve its goals. Therefore, it includes accountability which focuses on the attainment of the instructional outcomes and the improvement of the educational situation. These objectives are considered as a basis for the construction of each specific test.

1.8- The Different Steps of the Test Construction

The test goes through various steps before it obtains its final form. First, the tester has to think about the test objectives. Then, he chooses the tasks and writes them. After checking the suitability of the test content and the scoring system to the testing situation, the tester administers the test. Carroll (1980) claims that test construction involves three phases referred to as design, development and operation.

Design deals with the specification of the content of the test according to the objectives wanted to be achieved from it. There must be a clear understanding of the learners' abilities and their perception of the learning situation so that to be able to set specific tasks that will help the tester to get the results he wants for the sake of making decisions.

In the phase of development, the tester writes the test and analyses its different items. He also assigns specific scores to the test items and finds out if there will be no difficulty in the process of scoring.

The test operation involves the administration of the test which is put into practice to get information about the learners. Therefore, the test designer has first to think about what to test and how to test it. When the test objectives and tasks are specified, the test is prepared and its items are analysed. Finally, it is administered and scored.

"The most important consideration in designing and developing a language test is the use for which it is intended, so that the most important quality of a test is its usefulness" (Bachman and Palmer, 1996: 17). The test usefulness refers to the

degree to which a test is capable to achieve its purpose. It is an essential quality that is taken into account in the process of test design and development. Bachman and Palmer (1996) view test usefulness as an interrelated function of six test qualities that include reliability, construct validity, authenticity, interactiveness, impact and practicality. However, some researchers point out that authenticity, interactiveness and impact constitute a part of validity (McNamara and Roever, 2006). This means that the test usefulness requires reliability, validity and practicality.

The phase of test design and development is very important as it leads to the establishment of specific purposes for the test to make it useful when it is utilized. Before being administered, the items of the test are tested using a process called item analysis so that to get data about how the test will work.

1.9- Item Analysis

Item analysis refers to the process of checking the effectiveness of the different tasks involved in the test before administering it. It aims at asserting that the various test items are really suitable to the testing situation. *"This procedure involves the careful analysis of score patterns on each of the test items. The analysis tells us how well each item is working"* (McNamara, 2000: 32).

Item analysis takes place in the phase of test development in which the items of the test are tested to check whether the test is valid, reliable and scorable. It tries to gather information about the results of the test administration, the suitability of the test content to the abilities of the learners and the ease of scoring the test. The criterion testees for the analysis have to be carefully selected by focusing on their profiles to the clarity of this procedure (Carroll, 1980).

This technique can be used to analyse items in classrooms tests, questionnaires or other research instruments. It relies on the use of various methods to provide data about the nature of the elements of the test. Item difficulty and item discrimination are the most typical methods employed for analysing items.

Item difficulty is an index that indicates if the test items are easy or difficult to answer. It refers to the proportion of people getting an item correct. This proportion (*P*) is calculated by dividing the number of individuals answering correctly by the total number of the people responding to the question.

Thus, " $P = \dfrac{\textit{Number who answer correctly}}{\textit{Total number of individuals}}$ "   (Whiston, 2008: 83).

The item difficulty may range from 0.00 (no one answered correctly) to 1.00 (everyone answered correctly).

Item discrimination, or item discriminability, gives information about the extent to which a test item discriminates between the testees. It enables to differentiate between the performances of the learners. Item discrimination is calculated using several methods but the most common one is called the extreme group method which involves dividing the examinees into two groups based on high and low scores. The item discrimination index (*d*) is obtained by substracting the proportion of the testees in the lower group from the proportion of the examinees in the upper group. Therefore, " *d= upper% - lower%* " (Whiston,2008).

Item analysis takes place before a test is administered. It leads the test to be more acceptable since there will be a certain evidence that it will assess the learners' abilities in the right way and that the scorers will not find problems in scoring. The test items differ according to the testing purposes. In the assessment of writing, the tasks are related to the nature of the assessment methods.

1.10- The History of Writing Assessment

Testing is not a new phenomenon since it is claimed that it was employed by the Greeks around 2500 years ago and by the Chinese in a civil services examination around 2000 years ago. Nevertheless, the launching of the testing movement took place thanks to the English biologist, Francis Galton, who wanted to measure

biological individual characteristics in a study of human heredity during the late 1800's and early 1900's. Therefore, he utilized sensory discrimination tests to measure individual intelligence.

The process of assessment has been also influenced by the ideas of Alfred Binet and Théodore Simon who worked on intelligence tests. In 1905, a measure of intelligence called the Binet-Simon scale which focused on assessing comprehension and reasoning was introduced. It included thirty items. In 1908, it was altered to incorporate the intelligence quotient (IQ).

As intelligence tests were administered individually, researchers started thinking about group testing. During the First World War, military leaders in the United States of America thought about using new models of tests. In 1917, the army hired a group of psychologists who were guided by the American psychologist Robert Yerkes in order to develop group intelligence assessment which aimed at the selection and the classification of the personnel. The group administered intelligence tests used a multiple-choice format and were classified into two categories. The first category which was called Army Alpha was employed for routine testing. The second type which was referred to as Army Beta was used with illiterate or non English speaking recruits. After the First World War, the army released these tests for public use (Whiston, 2008).

Although testing has been around for many centuries, its methods changed from a period to the other. Concerning written examinations, it can be stated that they appeared in the middle of the nineteenth century but they became the most common method of assessment in the late nineteenth century and the beginning of the twentieth century.

Before the 1850's, examinations took the form of annual arbitrary recitations. In 1845, Boston grammar and writing schools introduced written examinations. However, the problem of accurately scoring students' writing was raised. In 1846, there was the introduction of a point system for determining proficiency in composition (Barone and Taylor, 2005).

In 1850, written examinations became a major tool of testing in England and emerged in USA. They were introduced in India in 1853. Between 1850 and 1870, the idea of making a link between education and examinations largely spread. Thus, standards of qualification in schools and universities developed (Roach, 1971).

In the 20th century, quantifiable approaches were created to evaluate writing. In 1903, a scaling system was introduced by Dr. Joseph Rice. This system used a five-point scale (100, 75, 50, 25 and 0) to rank learners' papers as excellent, good, fair, poor and failure. The development of a scientific testing movement called psychometrics led to think about various methods of assessing writing. One of the most famous writing evaluation systems was the Hillegas scale produced by Edward L. Thorndike and his student, George Hillegas in 1912.

In the mid 1930's, written examinations were known as the new type test composed of objective questions. During this period, there was an attempt to formulate essay questions that could be marked objectively. However, objectivity meant to alter the form and function of the essay test (Ruth and Murphy, 1988).

From 1950 to 1970, assessment of composition relied on measuring writing ability by making the learners answer questions about grammar, usage and punctuation in a multiple choice test. Objective testing became popular. However, the problem was that it was reliable but it was not valid since the students' scores did not reflect their writing abilities (Lippman, 2003).

Some educational researchers collected a set of ranked composition examples on the basis of form and content. The teachers had to match their students' writing with these graded writing samples so that a uniform grade was given. These types of scales for comparison continued to be popular until the 1970's. During the 1960's, direct writing assessments were used. They required the students to write in response to an assigned topic or prompt. The papers were scored on the basis of a set of criteria (Barone and Taylor, 2005).

From 1970 to 1986, writing assessment focused on essay tests that were written in a single session. These types of tests were considered as being more valid than objective tests since they measured writing proficiency rather than grammar proficiency. From 1986 to the present, testing writing involved the

evaluation of essays written at a given period of time or the assessment of portfolios. Over the years, students' writing started to be assessed through alternative objective and holistic methods. Some universities use objective tests while other universities employ the holistically scored essay examination (Lippman, 2003).

The assessment of writing may involve a variety of tasks.

1.11- The Types of Tasks Used for Assessing Writing

The assessment tasks refer to the types of questions or activities used to test the learners' abilities. They are related to the kind of skill or area of knowledge designed to be tested. *"It is important to select testing methods (tasks) that , on the one hand , accurately reflect the kinds of language skills you want to test"* (Genesee and Upshur,1996 : 154).

Therefore, the test objectives determine the type of testing tasks to be used. Assessment tasks may involve three types of formats which are open-ended, closed-ended and limited-response formats.

The open-ended response format includes tasks that do not require limited answers since the testees are free to express themselves in whatever manner they want. They are scored in a subjective way since the scorer cannot predict the test answers and design definite marks for them. Interviews and essay writing are examples of these tasks.

The closed-ended response format is based on the use of multiple-choice or selected response items. The learners have to choose the correct answer from a set of alternatives. They are utilized to assess the learners' comprehension of listening or reading passages. They require an objective scoring as the responses have to be considered as either right or wrong.

The limited-response format involves tasks requiring a limited number of answers which are not provided by the examiner but they are given by the learners.

They include cloze tests, matching or rearranging sentences and reproducing sentences stated by the teacher. The scoring of these types of tasks is subjective since there is not a single correct answer and the scorer has to rely on his own judgement.

Although the assessment of writing may involve various kinds of questions, two main types of tasks are distinguished. They are the convergent writing and the divergent writing tasks.

The convergent writing tasks, which are also referred to as constructed response items, focus on a topic-centered answer. They are used to measure the learner's content knowledge. They are often in the form of a limited-response format.

> *In convergent writing tasks, writers are expected to provide similar responses to items. They may range from dictation, with no choices in what should appear on their papers, to somewhat more flexible responses that call for logical inferences based on provided information or on content knowledge.* (Barone and Taylor, 2005: 106)

These types of tasks urge the learners to write to a particular test item which asks for a specific information. They are based on a set of expected right answers. Thus, the relevance of the answers is given more importance than the general features of writing.

The divergent writing tasks aim at measuring the learners' abilities to produce a written text. They include an open-ended response format. The learners may be required to write about a given topic suggested by the teacher. Sometimes, they may be asked to choose their own topics. Divergent writing tasks give importance to the form and the content of the written product (Barone and Taylor, 2005).

The differences in writing tasks imply that various methods can be used for testing writing.

1.12- Methods of Writing Assessment

The methods of writing assessment are classified into two categories referred to as direct and indirect methods. The indirect ones, which are also called objective writing assessment methods, often include grammar and spelling tests. *"Objective writing assessments measure discrete aspects of writing such as verb tense, sentence structure, grammatical usage and vocabulary in a multiple-choice format"* (Barone and Taylor, 2005: 99).

They employ objective questions that are scored by the assessors in a mechanical way. These methods are also related to automated scoring which is based on computer programs that count the number of spelling and grammar errors. They are seen as inefficient tools of assessing the learners' performance. *"Most writing instructors believe indirect methods provide only a crude approximation of the mental capabilities necessary for good writing"* (Finn, 1997: 171).

Since this type of assessment does not give a total description of the learners' performance, educators and writing instructors prefer to utilize the direct methods to test the students' writing abilities.

The direct methods of writing assessment put the emphasis on the learners' actual writing production. They employ holistic and trait-based scoring. Direct assessment is also called performance assessment. It gives importance to the measurement of the learners' ability by making them write an essay to lead them to demonstrate their writing competence rather than recognizing the correct answer (Lippman, 2003).

Weigle (2002) prefers to refer to this type of approach as the timed impromptu writing test. The direct methods of assessment give more freedom to the students in order to produce a written response that is evaluated on the basis of the rater's subjective judgement. The different methods of assessment rely on various types of scoring.

1.13- Scoring

Scoring is also called rating. It is the act of assigning a mark to the learner after correcting the test answer and making a judgement concerning the learner's performance. It often involves the scorer's opinion about the learners' abilities. It can be subjective or objective. Scoring is objective when the scorer gives grades to the learners according to a designed marking scale that considers each test item as being either right or wrong and rejects any other possibilities. For example, the rater makes an objective judgement when correcting exercises that deal with multiple choice answers or filling the gaps with one definite word.

Scoring is subjective if the scorer has to give his personal judgement. This means that the mark is based on the rater's appreciation of what he is correcting since there is not a definite answer and various responses are acceptable. For example, rating essays or conversations is most of the time subjective. There are three types of scoring: analytic scoring, holistic scoring and trait-based scoring.

1.13.1- Analytic Scoring

Analytic scoring involves the use of a marking scale that is based on several marks assigned to the different elements in the test. The combination of these marks gives the general score.

> *Analytic scoring is a form of assessment frequently used in the evaluation of writing, where a separate award is given for each of several nominated performance features of a particular task, for example relevance and adequacy of content, organization, lexical breadth and depth, in contrast to awarding one global score.* (Shaw and Weir, 2007:151)

This type of rating aims at testing the learners by awarding separate scores to various responses of the students. Analytic scoring provides useful feedback to the

learners and information to the teachers about distinct levels of performance that may be satisfactory or unsatisfactory (Genesee and Upshur, 1996). The use of analytic rating for scoring compositions or essays obliges the rater to take into consideration several features of writing. In this way the learner's ability to write an essay is judged according to his capacity to utilize various elements of the language including mechanics, vocabulary and grammar.

This approach supplies the scorers with diagnostic information through the assessment of separate features of writing by employing explicit criteria which can help the teachers to know about the learners' strengths and weaknesses in order to provide them with feedback concerning specific areas of their performance (Hyland,2003; Shaw and Weir,2007).

On the other hand, analytic scoring is time consuming and may cause a diversion in the purpose of essay assessment as it does not consider writing as a unity and regards it as a sum of distinct parts (Hyland, 2003).

1.13.2- Holistic Scoring

Holistic scoring involves a general judgement that is based on a single score mark which is awarded to the learner concerning his performance as a whole. It can be classified into types which are general impression scoring and focused holistic scoring. General impression scoring involves a global judgement of the students' assignment without focusing on a set of writing criteria (Finn, 1997). Thus, the scorer does not split the answer into parts but he corrects it in general. Since there are not different grades designed to specific elements in the test only one mark is given according to the scorer's subjective judgement. Focused holistic scoring requires the raters to test the learners' pieces of writing by assessing different aspects of the written product and combining these evaluations in an overall impression formulated in one score (Barone and Taylor, 2005). In general, these two types of holistic scoring rely on the scorer's impression.

> *Holistic scoring, often referred to as impressionistic marking, involves rating scripts impressionistically on a single rating scale according to their overall properties rather than providing separate scores on specified features of the language produced.* (Shaw and Weir, 2007: 150)

It is concerned with the overall performance of the learners without giving importance to specific abilities. The given score is assigned on the basis of the scorer's opinion about the quality of the piece of writing.

The assessor is more interested in the strengths of the written product. Thus, he does not emphasize on all the types of errors produced by the testees. The holistic approach to writing assessment considers writing as a single entity that is rated on the basis of a single scale integrating the different qualities of the writing sample (Hyland, 2003). The main advantages of this type of scoring are the emphasis on achievement and the rapid way of assessing compositions which makes their marking economical (Shaw and Weir, 2007).

Holistic scoring is considered as an efficient approach to writing assessment (Ferris and Hedgcock, 2004). The disadvantage of this kind of rating is the fact that it does not provide appropriate feedback and diagnostic information since the scorers award a mark on the basis of a global impression without distinguishing between the distinct features of writing (Shaw and Weir, 2007).

Generally speaking, holistic scoring requires a subjective view that relies on the rater's degree of satisfaction of the learner's response.

1.13.3- Trait-Based Scoring

Trait-based scoring is an approach of assessing writing that requires particular criteria to each prompt. It is context-sensitive since it makes a judgment of performance traits that are linked to a special task (Weigle, 2002; Hyland, 2003). In fact, the trait refers to a specific feature of writing. For each assignment, the distinct genre features of the text that should be produced are defined. This means that a

trait-based scoring guide is individualized to every prompt and rubric preparation is incorporated into the process of task design (Ferris and Hedgcock, 2004). This approach is distinguished by the employment of two main techniques referred to as primary-trait scoring and multiple-trait scoring.

Primary-trait scoring represents a form of holistic rating that focuses on the assessment of a piece of writing on the basis of one critical feature of the given task such as effective arguments, creative response or reference to sources (Hyland, 2003). The written product is awarded a score with reference to the principal trait that is criterial in the performance of specific rhetorical assignments (Shaw and Weir, 2007). The scoring guides of this approach are based on the types of writing (narrative, expository, descriptive or argumentative) as well as the audience, the purpose of writing and the kind of task involved in the process of assessment (Barone and Taylor, 2005). Primary-trait scoring is generally employed in research situations or in a teaching course to get data about learners' mastery of specific writing features (Shaw and Weir, 2007).

Multiple-trait scoring assigns separate scores for several traits in a single piece of writing. Some authors consider it as synonymous to analytic scoring since it focuses on different writing features (Shaw and Weir, 2007). However, the former may differ from the latter which insists on the established criteria relevant to a specific task.

> *Multiple-trait scoring treats writing as a multifaceted construct which is situated in particular contexts and purposes, so scoring rubrics can address traits that do not occur in more general analytic scales. These might include the ability to 'summarize a course text', 'consider both sides of an argument', or 'develop the move structure of an abstract'.* (Hyland, 2003: 230)

This approach may focus on different traits such as ideas, details, purpose, organization, vocabulary, syntax, punctuation and spelling. Some traits are occasionally given more importance than others (Barone and Taylor, 2005). This technique can provide information that can be used for remedial work.

Primary-trait scoring is based on one feature while multiple trait scoring relies on several traits. However, these two systems have the same purpose which is the judgement of the quality of the piece of writing in relation to a specific context. The advantage of trait-based scoring is the precision of the trait-based descriptors which help the assessors to resolve discrepancies in a better way than holistic and analytic scoring. However, this procedure has certain disadvantages as the construction of a scoring guide for each writing assignment is time consuming.

When using primary trait scoring, the raters may find a difficulty in focusing on one specific trait and diagnosing the learners' weaknesses. In addition to this, both primary and multiple trait systems may lead to subjectivity due to the problem of the identification of a single facet that is appropriate to the task (Ferris and Hedgcok, 2004).

In order to correct the test answers and award the testees specific marks, the scorer needs certain rating scales that help him in scoring the test.

1.14- Types of Rating Scales

A rating scale consists of a number of marks designed for the test items. It is a means of measurement that facilitates the scorer's task. It is defined as a framework consisting of a series of constructed levels of language proficiency against which a learner's performance is judged (Priestley, 1982).

The scoring procedure involves the use of a rating scale that formulates a set of qualitative criteria into a specific mark. *"A rating scale is a series of short descriptions of different levels of language ability. Its purpose is to describe briefly what the typical learners at each level can do"* (Underhill, 1987: 98). It helps the

assessor to give a score to each learner on the basis of the bands or levels to the descriptions provided in the scale. These descriptions are referred to as band descriptors or level descriptors which facilitate the interpretation of the rating scale (Lynch, 2003).

Various kinds of scales exist. Nevertheless, writing assessment is characterized by four types of scales which include analytic scales, holistic scales, primary trait scales and multiple trait scales (Weigle, 2002).

The analytic scale is employed for grading multiple tasks. It assigns a score for each component which means that it analyses autonomous text features (Ferris and Hedgcock, 2004). It tends *"...to focus on superficial aspects of writing that could be quantified (for example, the number of spelling and grammar errors)"* (Lynch, 2003: 53). The analytic scale provides a numerical value to separate features of writing including content, organization, grammar, vocabulary and mechanics (Hyland, 2003).

The holistic scale gives a single score to the essay on the basis of scoring criteria that are in the form of descriptions of the syntactic and rhetorical qualities related to the writing performance (Weigle, 2002). In general impression scoring, the holistic scale requires the rater to provide a general evaluation of the writers' mastery of the different features of writing. In focused holistic scoring, the scale is assisted by a scoring guide or rubric which takes some elements from the analytic scale. *"Holistic scoring rubrics may comprise 4 to 10 levels or bands, each corresponding to a score, a set of descriptors, and 'benchmark' writing samples"* (Ferris and Hedgcock, 2004: 308). The descriptors used in these rubrics may be general or detailed.

A primary-trait scale provides a single score to one feature of writing. It involves the development and use of a set of guidelines for each prompt.

> *The guidelines comprise (a) the task (b) the statement of the*
> *primary rhetorical trait to be elicited (c) an interpretation of*
> *the task hypothesising writing performance to be expected (d)*
> *an explanation of how the task and the primary trait are*
> *related (e) a scoring guide (f) sample papers, and (g) an*
> *explanation of scores on sample papers.*
>
> (Shaw and Weir, 2007: 149)

This type of scales requires a detailed scoring guide that is employed to score those kinds of written products that involve in depth information linked to the evaluation of the mastery of a specific trait of writing or a particular content area knowledge (Barone and Taylor, 2005).

A multiple-trait scale, which is also referred to as an analytic trait scoring, awards separate scores to different traits. Thus, *"each task can be related to its own scale with scoring adapted to the context, purpose and genre of the elicited writing"* (Shaw and Weir, 2007: 230). The multiple-trait scale often relies on the use of the six traits of writing assessment. These traits are: ideas, organization, voice, word choice, sentence fluency and conventions (Arter and McTighe, 2000).

Among the rating scales discussed above, analytic and holistic scales are the most common ones that are often used in the field of language testing. Shaw and Weir (2007) claim that rating scales can be classified into two categories. The first category includes analytic scales, holistic scales, primary-trait scales and multiple-trait scales which refer to performance criteria. The second category comprises three kinds of scales that are identified by their function. These types involve the user-oriented, the assessor-oriented and the constructed-oriented scales.

User oriented scales are employed to know what the learners can do outside the educational system. They aim at providing information about the future behaviour of the students, at any given level, in work or social life. These scales are often holistic as they focus on one descriptor.

Assessor-oriented scales put the emphasis on the quality of the performance of the learners at each level. Some of these scales are holistic as they use one descriptor. Other assessor-oriented scales are analytic since they focus on various aspects of the learners' performance. Constructor-oriented scales are in the form of lists of specifications that concentrate on the kinds of tasks that the learners can perform at each level. They are holistic; they aim at guiding the construction of tests for each specific level.

Generally speaking, different types of scales exist. Wallace (1991) differentiates between numerical and letter scales. The numerical ones involve marks based on numbers. For example, some numerical scales involve a marking from 0 to 5 points. Others include a rating from 0 to 10 points while other scales go from 0 to 20 or 40 points. Each single score a learner obtains from the combination of the various points can be reformulated in the form of a percentage according to which a certain grade is awarded. The grades are transformed into letters to produce a letter scale in which the percentage of the learners can be represented as grade A, grade B, grade C and grade D.

Most of the marking scales used by the testers rely on a set of scores assigned to the test items. They vary from one to another according to the skills or elements of language tested and the number of items used. The more the rating scale is enlarged and composed of a number of scores, the more the scoring is objective because in this case the scorer does not need to give his personal judgement.

Testing the learners' mastery of language should include their performance in writing and speaking. However, the written and the oral language are assessed differently.

1.15- Testing the Spoken and the Written Language

In the assessment of writing the tester is concerned with the learner's accuracy and how well he can express himself without making too much mistakes. Accuracy is often linked to the mastery of grammar (Genesee and Upshur, 1996).

Thus, the testee's answer has to be linguistically correct. In addition to this, the scorer must focus on assessing the degree of appropriateness of the responses to the given question. This means that both form and content are taken into consideration when testing the learners' writing performance. Also, the quantity or amount of information provided by the testee is given importance because it shows the learners' ability to use the target language to express himself and convince the reader.

The measurement of the learner's competence in using the spoken language creates more difficulties to the tester than assessing the written form since speech is evasive. *"The testing of speaking is widely regarded as the most challenging of all language tests to prepare, administer and score"* (Madsen , 1983 : 147).

When assessing the spoken language, the assessor should be concerned with the learner's fluency which is related to the degree of speed with which the learner delivers his speech and the amount of information he can provide during a certain period of time without having too many interruptions and pauses.

Another element which is tested in oral performance is accuracy which refers to the correctness of the sentences uttered by the testee. In addition to this, the tester assesses the learner's pronunciation which is very essential because it is related to the degree of comprehensibility of the testee's speech. Thus, *"The real test of whether he can speak the language is whether he can say something which is understood by the listener as relevant to a particular situation"* (Perren, 1968: 112).

The testing of speaking focuses on the learners' ability to express themselves orally in an intelligible way. Their speech must be comprehensible and appropriate to the content of the test.

The learners' writing performance is assessed in a different way from that one which is related to the use of the spoken language since the tests' tasks and methods of assessment used for the former differ from those involved in the latter.

However, the assessment of these two skills focuses on the examinees' accuracy, their mastery of vocabulary, the appropriateness of their answers and their ability to express themselves.

Testing is a very important process that helps to provide information about the learning and teaching situation but sometimes certain difficulties are encountered by the testers when administering and scoring tests.

1.16 - Some Problems Related to Assessment

The testing process can face several problems, among them; one can list invalidity, unreliability, the problem of authenticity, the problem of learners' individual differences and the problem of sampling.

1.16.1- Invalidity

Sometimes, a test is used to assess certain skills or abilities but its content does not reflect these objectives. It means that what is tested is not relevant to the aims and kind of decisions to be made from it. The test is not valid if it is not constructed according to its use. In this case, the results obtained will not reflect the real abilities wanted to be tested. For example, if a test aims at assessing the learner's ability to use the language orally and the question involves writing a dialogue, it cannot be considered as being valid since the content of the test will assess a different ability from the one which was supposed to be tested. As a result, the tester will not be able to make the right decision and there will be a deviation from the purposes designed for the test.

The invalidity of the test will hinder it from reaching its objectives and the examiner will find a difficulty in deciding whether the learners' performances reflect their mastery of the elements which were supposed to be tested or not. In addition to this, an invalid test will have a negative impact on the learner's motivation, *"A testee's performance is much different when he knows, or suspects, that the test has no relevance to his future"* (Carrol, 1980: 69).

Therefore, the test will not assess the learners' abilities in the right way when it is not valid. This is why the test designer has to make sure that the test is relevant to the objectives it is set for.

1.16.2- Unreliability

One of the problems that may be faced when testing the learners is unreliability since a test is considered as being unreliable if the results obtained from the administration of this same test at different occasions are not consistent.

Several factors may affect the test's reliability and lead the scores to fluctuate from a situation to another. For instance, the students' scores may change because of the lack of the learners' motivation or the unclear instructions. The learners' personal problems can also affect the test results too. Thus, the scorers cannot predict the reliability of the test in future occasions because there are certain external conditions such as the learners' motivation or administering the test in an inappropriate way which may cause the scores to diverge from the ones obtained before. This means that the measurement of the learners' abilities can be affected by certain unpredictable factors.

In order to try to make the test possess reliability, the testers should be concerned not only by the content of the test and the way of scoring it but also by providing good conditions for administering the test. However, the learners' motivation cannot be totally controlled and it will always have an impact on their results.

1.16.3- The Problem of Authenticity

In order to be efficient, a test has to assess the learners' ability to use the language for communicative purposes. If it does not take into account the fact that what is tested resembles certain concrete situations in which the learner must use the target language, the test is supposed to lack authenticity.

> *The most complex and persistent problems in language testing are those presented by the consideration of the relationship between the language use required by tasks on language tests and that which is part of our everyday communication use of language.* (Bachman, 1990: 356)

When the test does not provide information about the learners' ability to use the language in real-life, one can say that there is no authenticity between the kind of language use required in the test and the one which the learners need in authentic situations. For instance, if a group of learners are attending a certain course in order to use a foreign language for making scientific studies and they are tested on how to analyse poetry or literary texts, the testing process will lack authenticity.

Language teaching and testing are interrelated. The former aims at making the learners master the language for the purpose of using it in real-life situations while the latter is concerned with the degree of the effectiveness of the teaching process. This means that testing gathers data about whether the learners are able to use what was taught to them in order to communicate or not. In order to achieve the real purposes of assessment and make it provide sufficient information about the learners' abilities, the test should be constructed in a way that approximately resembles authentic tasks and real-life situations.

1.16.4- Individual Differences

The test designers find some difficulties when constructing tests because they have to take into consideration the learners' individual differences. The aim of assessment is to diagnose the learners' abilities and to find the degree of differences in their performances. *"Tests should be designed to reveal not simply the number of items which are answered correctly, but to reveal the quality of the candidate's language performance"* (Morrow, 1979: 145).

The learners differ in their level of intelligence, their pace of learning, their motivation and their way of conceiving and answering the test questions. This is why, the design of tests is a difficult task since the one who writes them has to provide items that are neither easy for some learners nor too difficult for others in the sense that the test items have to be considered as being acceptable by all the learners who are very distinct from each other. However, *"it is extremely difficult to develop tests in which all the tasks or items are at the exact level of difficulty appropriate for the individuals being tested"* (Bachman, 1990: 37).

Individual differences create some problems for the test designers as they take more time in trying to find which elements will best suit the levels of all the learners. Therefore, the test items should be answerable by a large number of the group of testees.

1.16.5- The Problem of Sampling

When constructing a test, the tester may find a difficulty in determining what items should be tested. The test designer has to keep in mind that the items used to measure the learners' ability have to be clearly established so that to get precise information about their knowledge concerning a certain specific skill or element of the language studied.

> *A test which only set out to measure the learner's semantic knowledge would not sample his whole knowledge and would therefore be invalid … The problem of sampling is thus closely bound up with the adequacy of the linguistic model or 'construct' chosen.*   (Corder, 1973: 369)

Thus, the way in which the test is constructed and the nature of the selected items will determine the effectiveness of the testing process. The tester is always doubting and hesitating about what to include in the test questions and what to reject so that to ascertain that the learner's competence concerning a certain area of knowledge is really measured appropriately.

The items designed to be used in the test constitute a sample of the elements needed by the learners to use the language efficiently. They will help to measure the learner's ability to perform well. In addition to this, this kind of sampling will determine the degree of the validity of the test.

However, testers find difficulties in deciding about the kind of activities or tasks they will use to test their students' knowledge since they are limited by time and they cannot predict the degree of their learners' understanding. Also, the content of

the test cannot be identical to the whole content of the studied courses and will not give a clear idea about the general capacities of those students. The sample chosen will measure the learners' competence as far as the items included in the test are concerned.

1.17- Conclusion

This chapter has focused on the description of the main concepts related to testing and assessment. It has provided some definitions of testing, assessment and evaluation. It has dealt with the different types of tests as well as the characteristics of a good test including validity, reliability, discrimination and practicality. Then, methods and purposes of assessment, the different types of scoring, rating scales and tests' effects have been tackled. In addition to this, a brief explanation has been provided concerning the principal elements assessed when dealing with the testing of the speaking and writing skills. Finally, some problems related to assessment have been discussed.

CHAPTER TWO:

Teaching and Testing English at Higher Level with Special Reference to the University of Tlemcen

CHAPTER TWO:

Teaching and Testing English at Higher Level
with Special Reference to the University of Tlemcen

2.1- Introduction

There is a close relationship between teaching and assessment which seem to be overlapping as both of them are complementary to each other. This second chapter provides information about the teaching and testing situation at the Algerian university with special reference to the University of Tlemcen.

First, it speaks about the concepts of teaching and learning. It also states certain considerations related to teaching English as a foreign language as well as teaching writing to EFL learners. Then, it deals with teaching English at the Algerian university involving mainly its purposes and the modules taught to students preparing a degree of English. In addition to this, the broad lines of the general programme of teaching English at the University of Abou Bakr Belkaid at Tlemcen are presented. Finally, a description of the way of testing English at university is provided.

2.2- Teaching English as a Foreign Language

The English language evolved throughout the centuries to reach the status of a global language. The study of the historical development of English led to the distinction of three periods within which this language has progressed. The language used in the period starting from the fifth to the eleventh century is called the Anglo-Saxon or Old English. About a third of the language employed today belongs to Old English which utilized the runic alphabet consisting of twenty-four letters. Middle English refers to the language of the period stretching from the twelfth century to the middle of the fifteenth century. It was characterized by the influence of French due to the Norman conquest of 1066. It has also borrowed vocabulary from Latin during the fourteenth and fifteenth centuries. It employed the French orthography instead of that of Old English. Sometimes, Middle English is called the language of Chaucer. In this era, there was a change in the pronunciation of vowels.

institutes in their own country or as transitory visitors in a target-language country" (Harmer, 2004: 39).

This means that the activity of teaching English to this category of learners has to take into considerations their linguistic background and their educational needs.

Teaching a language implies the application of teaching methods that focus on making the learners able to use the language for communicative purposes. *"The learner's effectiveness in communication relies on his mastery of the formal patterns, or usage, of the language"* (Carroll, 1980: 7).

Thus, teaching a foreign language should help the learners to master the target language so that to be able to express themselves both orally and in writing. It has to make them equipped with the necessary elements needed for communicating via the language. The most essential language elements are grammatical structures and vocabulary.

The effectiveness of the teaching of a foreign language lies in the types of methods and techniques used. The teacher has to know certain strategies which will lead him to be aware of how to transmit information to the learners and make them motivated to receive and memorize them. In addition to this, the teacher must be conscious about the changes occurring in the target language since *"…living languages are always changing structures"* (Spolsky, 1998: 78).

The teachers of EFL (English as a foreign language) may encounter certain problems in the teaching process due to the difficulties faced by the learners in their process of learning.

A learner of a foreign language learns in a different way from the native speaker. The latter learns the language from infancy and is not aware of the process that is taking place as it is part of his daily life. His learning of the language is imposed on him by his need to communicate with his family and society. This means that the acquisition of the first language goes hand in hand with the person's discovery of the people and environment that surround him.

The learner of English as a foreign language cannot learn this language in the same way as his mother tongue because the latter is acquired and is already consisting a means of communication while the former is a new system that is not very often used by those who interact with him. When learning a foreign language, the process of language transfer constitutes a serious problem for the foreign language learner who tries to employ this first language to translate his thought. This leads him to make mistakes because of the differences existing in the structures of both languages. Also, the learners may find difficulties in practising this language since it is not often used outside the classroom.

Learning English as a foreign language requires more concentration and effort than the acquisition of the mother tongue. *"The consciously learned system which concentrates on form cannot initiate utterances but can only serve to monitor (correct, revise) them. The learner must acquire enough of the target language to initiate utterances"* (Hawkins, 1984: 93).

Therefore, the learner of English as a foreign language needs not only to learn the structures of this language but also he must get a considerable stock of vocabulary so that to be able to produce sentences. In order to achieve this stage, much effort and time has to be spent. The mastery of both grammar and vocabulary will lead the learner to communicate efficiently.

The teaching of English as a foreign language is a difficult task mainly because of three complex factors that may make this activity a very complicated one. First, the learners are not exposed to this language outside the classroom. Second, they do not have the opportunity to employ it in real contexts. Finally, the majority of EFL teachers are non-native speakers of English whose job is not only to teach this language but also to acquire a sense of self-confidence to use English in a way that approximately resembles the situations encountered in the communities of the target language. Therefore, the main role of the teacher is to make the learners able to communicate providing them with sufficient information about the target language (Kramsh, 1993).

Teaching a foreign language should lead the learners to achieve a certain level of performance that enables them to make communication inside and outside the classroom (Littlewood, 1984). Thus, in EFL settings, *"...the burden for providing the cultural dimension to the curriculum very much rests with the teacher"* (Carter and Nunan, 2001: 2).

The teachers of English as a foreign language must focus on the authenticity of the language use. He has to be aware of the learners' abilities and degree of competence so that to decide about the content of the lessons that will suit their levels outside the classroom.

In such a context, the learners totally rely on the teacher in order to master the language. The latter is responsible for the kind of instruction the former obtain.

When teaching English as a foreign language, both the speaking and writing skills are given importance. However, the teaching and testing purposes often diverge the teachers' attention towards the writing skill.

2.3- Teaching Writing to EFL Learners

The purposes of teaching writing differ from a specific context to the other. When teaching writing to the learners who study English for specific purposes, the writing tasks focus on developing the learners' language competence especially in the domain of their study. Thus, they should be trained to produce pieces of writing that are relevant to their learning situation and their specific needs.

On the other hand, teaching this skill to the learners of English as a foreign language is based on different purposes. The teacher gives more importance to the improvement of the students' general language competence. The writing activities have to be suitable to the general needs of the learners. *"In the case of adult students, it is often hard to find writing tasks that are directly relevant to the varying needs of a class full of students from different backgrounds"* (Harmer, 2004: 39).

This means that the process of writing has to be linked to a particular purpose in order to make a decision about the nature of the topics and activities that should be employed to develop the learners' capacities.

Writing is a skill that requires sufficient knowledge about the language to produce a written text. It is a difficult task as it involves cognitive processes. When teaching this skill to learners of English as a foreign language, more efforts and techniques are needed since these students often find problems in writing well.

> *The process of text generation, or encoding internal representations (ideas) into written text, may be disrupted by the need for lengthy searches for appropriate lexical and syntactic choices. Consequently, the written product may not match the writer's original intention.* (Weigle, 2002: 36)

Therefore, some learners of English as a foreign language may encounter certain difficulties in writing compositions because of their limited linguistic knowledge. In addition to this, the lack of motivation, anxiety, and the cultural differences existing between the mother tongue and the target language may have great effects on the learners' writing performance.

According to Hartley (2008), research about writing processes has proved that the production of essays or written texts involves a hierarchy of overlapping levels. The first level refers to writing as a social process since there is a relationship between the purposes of writing and the way of producing texts. The second one is concerned with writing as a cognitive ability since a composition or a written text is a reformulation of thought. Finally, the process of putting pen to paper is classified at the bottom level of the processes of writing.

The teacher of English as a foreign language has to take these levels into consideration in order to make the learners understand how writing functions.

Another important issue is related to writing as a process versus writing as a product. Process writing links this skill to the thinking process. It puts emphasis on how to produce and link ideas. It focuses on training learners to become creative writers (Hyland, 2003). Product writing focuses on the finished written text. It gives more importance to the final construction of the text and the language employed by the learners. In the past, the teaching of writing emphasised on the written product more than the writing process. Recently, the concept of writing as a process has also been given importance (Harmer, 2004).

Various approaches are employed to teach writing including product, process and the genre-based approaches. The product approach urges the learners to produce essays by focusing on what they have to write. The process approach is based on teaching the students the four stages of writing which are: planning, drafting, revising and editing. It helps to know about how to generate ideas and organize them to reach the final version of a composition. The genre-based approach aims at making the learners aware of the different ways employed to organize information in written texts by giving them the opportunity to practise writing essays involving different types of writing (Richards and Renandya, 2002).

The teachers of English as a foreign language have to use these approaches to teach writing according to the levels and the needs of their learners as well as the purposes of teaching this skill. The practise of these approaches takes place through a variety of writing tasks (Harmer, 2004). However, over-emphasis on one approach only may be insufficient for producing competent writers. The product approach can be beneficial for the students but it ignores the stages of the writing process. Even if the process approach encourages the learners to reflect on how written texts are produced, it is not easy to implement because it is time-consuming and does not enable the learners to do much writing as they spend more time in reflecting and rewriting. The genre-based approach helps the learners to understand the different purposes of writing by focusing on genres but it requires extensive practise. For all these reasons, it is more suitable to opt for a methodology of teaching writing that is

2.4- Teaching English at the Algerian Universities

English is taught as a foreign language in Algeria. At the middle and the secondary schools, this language is studied just as one subject-matter among others. It is not assigned a very high coefficient. Failure or success in learning English at these levels does not really affect the learners' degree of achievement in their studies as some of them may pass even if they did not perform well in English while it may happen that a learner will fail although he got good marks in the foreign language. This is due to the fact that the decision to be made about the pupils' success or failure is based on the combination of several subject-matters.

As a result , some learners do not give much importance to learning English because they try to work better in those subject matters they are more accustomed to such as Arabic , history and Islamic education or those subject-matters having a high coefficient like natural sciences, mathematics, and physics (for scientific streams) or Arabic and philosophy (for literary streams). Therefore, EFL learning tends to be neglected at the middle and the secondary levels because English is neither a medium of instruction nor a language of study.

On the other hand, the situation of English at university is different from the other levels. At the institute of languages and arts, it is taught for general purposes for the learners who intend to become specialized in this language in order to get a licence degree and become future teachers of English as a foreign language. In this case, the focus is on the use of general English.

All the four skills are taken into consideration. However, speaking and writing are given more importance. In this case, English is considered as a language of study.

At the institute of science and technology, this language is instructed for specific purposes for those students who are learning scientific subjects and need English as a medium of instruction to have access to special references or information provided in this language.   Thus, this language is taught as English for specific purposes (ESP). In fact, ESP is divided into various categories but it is

usually classified into two main subdivisions which are English for academic purposes (EAP) and English for occupational purposes (EOP). English for academic purposes refers to the instruction of this language to university students who need to get a practical command of English. Various academic disciplines like science, technology, law, medicine and business fall under EAP which puts the emphasis on teaching reading and writing (Hamp-Lyons, 2001). English for occupational purposes is related to teaching this language to workers who need English for professional objectives like doctors, business people or engineers. Business English is the largest branch of EOP. This category focuses on teaching English vocabulary and the skills needed in the job of this group of people (Dudley-Evans, 2001).

The purposes of teaching English are adjusted to the learners needs. ESP teaching is based on using translation and vocabulary learning to enable the learners to use this language to consult scientific texts written in English. Teaching English as a foreign language at the Algerian university aims at making the learners master this language in order to use it in the future for occupational purposes. In fact, all the students whether they are studying English for general purposes or for specific purposes need to master this language because nowadays certain job opportunities can be available in different fields ranging from the sphere of education to the domain of business and commerce. This means that the chief objective of teaching English to these learners is to enable them to achieve an acceptable level of competence that helps them not only to perform the role of future teachers or get a job in any field but also to communicate with other speakers of English.

The teaching of English for graduate students involves different modules that are taught to the learners to lead them to achieve not only the mastery of general English and to get informed about the culture of the countries of the target language.

2.5- Testing EFL Learners' Writing Performance

The process of testing EFL learners' writing performance often takes place for placement or certification purposes. In fact performance assessment is often referred to as authentic assessment since it attempts to get information about the degree of the ability of graduate students to cope with situations that are likely to confront them in the real world. However, the teachers may find a difficulty in assessing EFL learners who may not display the kind of performance required from them due to their lack of linguistic or content knowledge. *"Students are more likely to have developed competence in taking exams that tested their knowledge, comprehension, and analytical abilities, but not the application of their acquired knowledge to a real-life or simulated exercise"* (Finn, 1997: 166).

This means that performance assessment may be a difficult task both for the teacher and the students as it is linked to the measurement of the learners' ability to use what they have learnt in order to complete activities resembling the tasks that they may encounter in real life situations.

When the learners are compelled to write essays, they require more cognitive effort in order to produce pieces of writing whose content and structure are acceptable. *"The distinction between knowledge telling and knowledge transformation is a useful notion for both writing pedagogy and assessment"* (Weigle, 2002: 35).

The tasks that ask the students to retrieve the information given to them may not provide sufficient data about their writing abilities. Writing performance assessment should involve tasks that lead the learners to employ the acquired knowledge in order to reflect their understanding of the topic and discuss it in a logical way. Therefore, the content and degree of difficulty of the writing tasks constitute an important issue in the process of testing EFL learners.

Assessing the students' compositions is a complex process that requires more energy from the part of the teachers as they have to rely on their own impression and have to read carefully the learners' papers to provide an accurate judgement of the

students' performance. *"The evaluation of student writing continues to present teachers and administrators with a plethora of challenges"* (Lippman, 2003: 200). The assessment of writing is confronted by various problems. For instance, the issue of form and content is often debated. Thus, some assessors may focus on the form of the essay while others may give more importance to its content.

A good piece of writing should be both cohesive and coherent. Cohesion refers to the linguistic techniques including lexical and grammatical devices that bind the elements of a text together. Coherence means the internal logic of the text as well as information sequencing that enable the reader to understand the ideas stated. However, certain texts may be cohesive but not totally coherent while some other texts may be coherent but they are not cohesive (Harmer, 2004).

This kind of situation can create difficulties for the scorers who may be diverted from their task of assessing the learners' general performance to the act of testing either form or content. The issue of form often pervades in this process. *"Because grammatical errors are among the most obvious and salient characteristics of students writing for many readers, grammatical form may receive more attention from raters than other text features such as idea development and organization"* (Ferris and Hedgcock, 2004: 241). In the case of EFL learners, the teachers are more concerned with the students' ability to use the language rather than their amount of knowledge about the modules related to content areas. This is why they may ignore the fact that the learner did not excel in discussing the topic if he has a good style and reveals a considerable extent of the mastery of the English language.

Nevertheless, some educators and researchers put the emphasis on the content of the student's essay. *"Quality, not quantity, is the central feature of idea development. It is using precise information with specific details in just the right place(s) that make for an excellent piece of writing"* (Barone and Taylor, 2005: 104). Therefore, the students' way of developing and presenting ideas will reveal their degree of understanding and their ability to discuss and analyse the problematic raised in the examination question. Generally speaking, a good essay is the one

Generally speaking, testing the writing skill is a difficult process, when it is related to a context of EFL learners' assessment, this task becomes more complicated as the first language and culture of these learners will influence their way of thinking which will be reflected in their pieces of writing. Also, their limited linguistic knowledge will shape their way of answering the questions of the examination. This is why the teachers have to set up clear tasks methods for the evaluation of students' essays.

The tools used for assessment should be linked to the purposes of such a process. The learners have to be prepared in advance to cope with the testing situation by making them practise writing for a variety of audiences and purposes during the teaching sessions designed for this skill.

2.6- Teaching English at the Department of Foreign Languages at Tlemcen

The University of Abou Bakr Belkaid was established in 1974. The foreign languages especially French and English started to be taught in 1988 within the institute of Arabic. In 1991, the institute of foreign languages was established. The department of foreign languages, including the sections of French and English, was separated from this institute in 1997. The section of English stayed under the supervision of this department until 2008 when it became an autonomous establishment. The number of permanent teachers of English is seventeen (17). The students of English obtain a licence degree after four years of study. EFL students have to study a certain number of modules that aim at making them master the target language and have a specific knowledge concerning the countries whose first language is English.

The following table provides information about the different modules taught at the four levels of the graduation studies of English at the department of foreign languages at the University of Tlemcen.

The module of Written Expression aims at enabling the learners to develop their writing skill by teaching them the different writing strategies. It is taught to the students of the first, second and third year level. In the first and second year, it is devoted three sessions per week. In the third year, the students study this module in one session.

The purpose of the module of grammar is to supply the learners with knowledge about the grammatical structures so that to be able to produce correct and accurate sentences whether orally or in a written form. The graduate students study this module for three years. They learn grammar within three sessions per week.

The aim of teaching phonetics to the first, second and third year students is to enable them to know the sounds of the language and practise pronunciation in order to speak fluently and be capable of communicating orally using the target language either with native speakers or with the teacher. The time allotted to this discipline is one hour and a half per week.

The module of linguistics makes the students acquainted with the main concepts and definitions dealing with language and its study. It is taught to the learners in the four years of their university studies and it is assigned the period of one hour and a half per week.

The students learn British Civilization in the second and third year to get informed about the main events happening in Great Britain, the way of life of the British people and the Kinds of changes that occurred in the British society. The period of time of teaching this module per week is one hour and a half.

The module of American civilization aims at providing the learners with sufficient knowledge concerning the history of the United States of America. It is

included in the programme of the second and third year level. The students study it once a week during a lecture which lasts for one hour and a half.

African Civilization is taught to the fourth year students in one session per week. It deals with the third world countries and the main issues related to them.

The module of British literature which concerns the literary movements as well as the literary works of British authors is studied in the second and third year. It is devoted the period of one hour and a half per week.

American Literature is taught to the second, third and fourth year level. The period of time assigned to this module is one hour and a half per week. It aims at providing the learners with information concerning some famous American writers and their works. In the fourth year, British and American literatures are taught within one module referred to as literature.

The purpose of educational psychology is to give the learners some brief explanations and definitions concerning the main concepts related to the psychology of education. This module is studied by the third and fourth year students in one session per week.

Oral expression is taught to the learners in the first, second, and third year. It is meant to make the students speak and express themselves orally using the English language so that to be able to communicate fluently in future situations. The number of the sessions of this module ranges from three sessions in the first and the second year to one session in the third year.

TEFL (teaching English as a foreign language) is instructed to the fourth year level for one hour and a half per week in order to supply the students with a certain knowledge about the main concepts related to teaching English.

Third World literature aims at teaching the learners the literature of the third world countries involving mainly writings of African authors. It is studied by the students in the third and the fourth year.

At the level of the third year, this module is learnt in two sessions a week. The first session takes the form of a general course and the second one is a tutorial or directed work (T.D) that incorporates a practical lesson. The fourth year learners deal with this discipline in one session per week.

The academic year 2009-2010 witnessed the introduction of a new mode of studying characterized by the LMD system (Licence, Master, Doctorate) which has been incorporated in the section of English in addition to the classical system that is based on the licence degree. The teaching of English in the LMD system also attempts to develop the learners' competencies through the instruction of various teaching units. Each unit comprises one module or more.

The teaching cycle of the licence is undergone in three years composed of six semesters. In each semester, there are fundamental units in addition to methodological units concerning methods of research, transversal units dealing with computer science in the first three semesters and psycho pedagogy in the remaining semesters. As well as units of discovery related to teaching psychology in the third and fourth semester while the fifth and sixth semester, the learners are required to make a training or present a project. The fundamental units taught in the first four semesters are provided in the following table.

Semesters	The four semesters	The four semesters	1st and 2nd semester	3rd and 4th semester
Units	Description of the English language	English language competencies	English cultures	English cultures
Modules	- Grammar - Phonetics - Introduction to linguistics	- Techniques of oral production - Techniques of written production -Discourse comprehension	-Introduction to literary studies. - Anglo-Saxon culture and civilization	- English literatures - Anglo-Saxon civilizations - African civilization

Table 2.2 - The fundamental units taught during the four semesters of the LMD instruction.

Starting from the fifth semester, the students have to study one of two options: language sciences or literature and civilization. Each option involves specific fundamental units which are illustrated in the table below.

Options / Semester	Language sciences	Literature and civilization
Fifth and Sixth Semester	Description of language: - Linguistic theories - Phonology - morphology	Literary culture: - Literacy theories - Compared literature - General culture
	Applications in linguistics: - Psycholinguistics - Sociolinguistics - TEFL	English cultures: - English literatures - Anglo-Saxon civilizations - African civilization

Table 2.3 – The fundamental units involved in the fifth and sixth semester

The testing process in the LMD system is also based on examinations that may involve indirect methods in the units related to the language skills as well as direct methods in the units related to content areas. In addition to the use of examinations, a form of continuous assessment takes place through the evaluation of the learners' homework, projects and training.

Since this research work is concerned with the classical system of the licence degree, the next sections of this chapter will focus on the teaching programme and form of evaluation of this category.

2.7-The General Programme of Teaching English at the English Section

In reality, there is not a specific official syllabus. The students of English study this language according to a certain programme designed by the teachers on the basis of general broad lines that focus on teaching the learners certain specific skills and language elements during the four years.

The general programme of each module concerning the four years at the English section at the University of Tlemcen is classified as follows:

2.7.1- The Content of the Modules Referring to Language Skills

In oral expression, the students should practise the use of idiomatic expressions and pronunciation as well as making dialogues, plays and discussions. Reading comprehension is based on the study of written texts.

In written expression, the students have to be taught about tenses, types of sentences, punctuation, clauses, paragraph writing, summarizing, essay writing, writing short stories and letter writing.

The module of grammar focuses on the teaching of the sentence structure, word order, types of sentences, negative forms, interrogative forms, articles, tenses, the

passive voice, auxiliaries, pronouns, prepositions, elements of the sentence structure, modification structure, quantity and amount, comparison and degree, location and direction.

2.7.2- Linguistics and TEFL

The programme of teaching linguistics to the four years is divided into four main parts. The first describes general linguistics. Thus, the learners are introduced to this concept through lectures whose content deals with definitions of language, language types , language varieties and the aims of linguistics. They also get informed about the scientific study of language, the domains of linguistics, linguistics terminology and language families.

The second section which concerns European structural linguistics gives the learners information about the studies of Ferdinand de Saussure and his role in founding modern linguistics. The lectures also involve explanations of the terms language, langue and parole as well as synchronic, diachronic, syntagmatic and paradigmatic studies in addition to the sign, signifier, signified and semiology. Then, the students are taught about phonology morphology and American structuralism including the theories of Edward Sapir and Leonard Bloomfield.

The third part of the programme puts the emphasis on structural linguistics. It involves the approach of Harris to structuralism, phrase structure grammar, context free rules and generative phonology. There is also the study of transformational generative grammar including the generativity of grammar, competence/ performance, surface structure/deep structure and the syntactic component. The learners are taught about semantics too.

The last section deals with applied linguistics. It attempts to teach the learners linguistics in relation to sociolinguistics and psycholinguistics. In sociolinguistics, the learners are taught about the sociology of language, social groups versus institutions, bilingualism, diglossia, social classes and socio semantic codes.

In psycholinguistics, the students should be introduced to this field by getting information about its main concerns. They are taught about language and thought (including stages and theories of L1 and L2 acquisition), the perception and interpretation of speech in addition to psycholinguistic constraints on grammar.

The programme of the module of TEFL includes definitions concerning teaching English as a foreign language, the process of learning and teaching languages, learning strategies, teaching methods and teachers' training.

2.7.3- Phonetics

The teaching programme of phonetics includes four parts. The first part deals with an introduction to articulatory phonetics in which the learners are given information about the sounds of language as well as the description of the speech organs, vowels and consonants. The second section focuses on the word in connected speech by teaching linking, elision , assimilation and junction.

The third part concerns stress and intonation. Thus, the learners are taught about stress and rhythm, word-stress and intonation patterns. In addition to this, a detailed study of intonation patterns is provided. The students also learn about the domains of phonetics (acoustic phonetics, pathological phonology, auditory phonetics). The theoretical part is consolidated by various exercises of transcription.

The last part of this module puts the emphasis on pedagogical phonetics. In this section, the learners should understand the use of phonetics in the training of teachers by concentrating on the importance of phonetics in language teaching, on pronunciation and phonetic transcription as well as the difference between American and British pronunciation.

This part also gives importance to the teaching of the segmental and the supra-segmental features. Different types of drills are employed to fulfil this purpose.

Finally, the learners are provided with various methods of correcting the mistakes made in pronouncing pronouns.

2.7.4- Civilization

The programme of the module of British civilization involves five main parts. The first section deals with the situation of Britain from the origins to the middle of the 18th century. Thus, the learners are taught about the social legislation in the 16th century, religion, the new scientific spirit and the rise of the first British empire.

The second section introduces the birth and growth of industrial Britain. It emphasizes on teaching the students about the origins and causes of the industrial revolution, the agricultural revolution as well as political changes occurring in Great Britain.

The third part is related to the domestic and emperial history of Great Britain. It gives information to the learners about the trade unions, political parties, the age of imperialism and case studies involving India, Egypt and Ireland in addition to Britain in the first and the Second World War.

The fourth section which is about decolonization involves lectures about the development of the commonwealth and the declining world role of Britain since the war. The fifth part concerns the British political and social life. In this section, the learners should be taught about the British political system, the welfare state, education, the mass media, immigration and race relations.

The programme of the module of American civilization comprises five parts. The first section, which is entitled the age of discovery and settlements provides an outline of American geography as well as lectures about the original inhabitants of America, the colonization of the new world and life in the colonies. The second part focuses on the birth of the American nation by exposing the students to a critical approach concerning the history of the American nation through the description of the growth of discontent, the American revolution, some principles from the articles of confederation and the constitution, the political economy of slavery, the civil war and the reconstruction.

The third section deals with the emergence of modern America via the study of industrial America (involving Robber Baron and the rise of the labour movement), immigration , culture in industrialized America and America between the two world wars. The fourth part surveys the political institutions of USA. Thus, the students are taught about the three branches of the federal government, the states and the political parties.

The last section tackles the political and social life in USA by giving the learners information about American imperialism, the problems of American minorities, the mass media, mass culture, elitist culture and counter culture.

The programme of the module of African civilization focuses on teaching the learners about the partition of Africa, the European colonization to African countries, the colonial rule including case studies of Algeria, Egypt and Nigeria, the liberation movements, the history of slave trade in Africa, panafricanism and South Africa.

2.7.5- Literature

The programme of British literature is divided into five main parts. The first section is entitled general survey on English literature from Chaucer to Blake. It aims at providing the students with a general description of the English literature that is related to old English and Middle English as well as the Poetry and drama of the Elizabethan and Jacobean period. The learners also study of literary devices and techniques through the works of Milton, Defoe, Richardson, Fielding and William Blake.

The second part provides a global survey of the 19th century English literature by introducing the students to the literary devices and techniques using as a basis the poetry of Samuel Taylor Coleridge, William Wordsworth, John Keats, George Gordon Byron and Percy Bysshe Shelly. The learners study some novels including the works of Elizabeth Gaskell and Charles Dickens. The Victorian poetry is

presented taking as examples Robert Browning , Algernon Charles Swinburne, Christina Rossetti and Thomas Hardy.

The third section deals with the study of some novels of the 19th century. It focuses on teaching the characteristics of the novel in the 19th century by providing a general presentation of two main authors of this period and the analysis of their novels. This study should include two novels which are: Joseph Conrad's Heart of Darkness and George Orwell's Animal Farm.

The fourth part concerns the study of some major poets and playwrights of the 20 century. Here, the learners should be taught about the 20th century English literature dealing mainly with the works of William Butler Yeats, Phillip Larkin and Bernard Shaw.

The fifth section concentrates on the study of the Elizabethan and Jacobean theatre by giving information about the structure of plays taking as illustrations Shakespeare's Hamlet and Macbeth.

The programme of American literature includes three parts. The first one presents a global survey from the colonial to the romantic period. It aims at making the learners study literary devices through extracts from the writings of John Smith and William Bradford.

The second section takes the form of a global survey from the realist period to the beginning of drama. It introduces the rise of realism and naturalism through the study of the characteristics of these literary movements taking as a basis the analysis of extracts from William Dean Howells's The Rise of Silas Lapham, Mark Twain's Adventures of Huckleberry Finn, Stephen Crane's Maggie: A girl of the streets and Benjamin Franklin Norris's The Octopus: A California Story. Also, a description of the literature of the modern period and psycho-realism is provided through some illustrations from Sherwood Anderson's Death in the Woods and Thomas Wolfe's Look Homeward, Angel. American drama is taught to the learners taking as a basis for this study plays written by Eugene O'Neill and Tennessee Williams.

The third part involves the study of the 20th century American literature represented by the works of some authors who tackled specific themes such as regional and naturalistic fiction which are dealt with through the study of Richard Wright's Black Boy, James Baldwin's Go Tell it to the Mountain and John Ernest Steinbeck's The grapes of Wrath. The style of William Faulkner is studied taking his work, The Bear, as an example. The students should be taught about American writings dealing with fiction and reality through Ernest Hemingway's Green Hills of Africa , For whom the bell Tolls and A Farewell to Arms in addition to Francis Scott Fitzgerald's Tender is the Night.

The programme of the module of third World Literature is divided into four parts. The first section presents the historical background of the Third World literature. Therefore, the students are taught about the definition and the characteristics of the Third World, the factors that led to the emergence of this type of literature, the features of this literature and the main African writers.

The second part concentrates on the oral tradition, its origins, its characteristics and its role as well as the African poetry and the different types of poems. A set of poems are also studied and analysed including Léopold Sédar Senghor's 'Prayer to Masks'.

The third section deals with the African novels. Here, the learners should get information about the written African literature, the rise of the novel in Africa, the famous African novelists and the main themes involved in African novels. In addition to this, a study of an African novel is provided in the analysis of Chinua Achebe's Things Fall Apart.

The fourth part focuses on the African short story and its characteristics, the learners have to study some short stories involving Chinua Achebe's 'The Sacrificial Egg'.

2.8- The Detailed Programme of English Concerning the Four Levels

The first year students are encouraged to learn the four skills in addition to the study of grammar and phonetics. Most of the modules employ separate exercises that help the learners to develop various kinds of capacities including the mastery of pronunciation, grammatical structures, vocabulary and writing strategies. The module of oral expression makes the students practise using the spoken language. The learners are also introduced to the nature of language through the lectures provided in linguistics.

The following table give information about the programme of English concerning the first year level.

lectures / Modules	Lessons	Exercises
Reading comprehension	- Study of texts	Exercises including comprehension questions, vocabulary activities and paragraph writing
Grammar	- Parts of speech - Modals - Articles - Tenses	Exercises involving identifying the different parts of speech , supplying articles and writing the correct form of verbs
Written Expression	- Strategies of writing - Punctuation - Types of sentences - Paragraph writing	Exercises involving supplying punctuation , identifying types of sentences , writing sentences and writing paragraphs
Phonetics	- Sounds - The difference between sounds and letters - Vowels - Consonants - Classification of sounds	Exercises of pronunciation and transcription
Linguistics	- Definitions of language - Definitions of linguistics - Traditional language studies - An introduction to Modern linguistics	

Table 2.4 - The programme of the first year level

In the second year, new modules are introduced in addition to the old ones. The learners are not only provided with information about the four skills but also about new concepts in relation to the language studied so that to make them accustomed to the history, culture and literature of Britain and USA because they refer to the native countries of the target language.

The following table gives a description of the content of the modules taught to the second year level.

Modules	Lessons	Exercises or text analysis
Grammar	- Tenses, Prepositions, nouns - Adjectives, reported speech - Active and passive voice	Exercises about writing the correct form of verbs, supplying prepositions , formation of nouns and adjectives, transformation of sentences
Written Expression	- Punctuation and capitalization - Subordinate clauses - Writing letters and dialogues - Paragraph and essay writing.	Exercises including supplying punctuation and capitalization, identifying the type of subordinate clauses, writing sentences, writing letters, dialogues and paragraphs.
Phonetics	- The syllables - Stress - Intonation	Exercises of transcription, stress and intonation.
linguistics	- Language properties - Morphology - Generative grammar - Semantics	-Exercises involving morphological description, tree diagrams and semantic roles

British civilization	- Britain in the 16th century and the 17th century - The industrial revolution - The working movements - Social life in the 19th century - Britain and the two world wars - Political parties	
American civilization	- The discovery of America - The British settlements - The reasons of emigration - The colonial period - The American revolution	
British literature	- Genres of poems - Techniques of versification - Drama - The rise of the novel - Enlightenment	- William Shakespeare's Othello, Macbeth and the Tempest
American literature	- Colonial literature / Settlers' reports - Puritan literature - The literature of reason and revolution -The age of Romanticism	- Extracts from J.Smith's 'The General History of Virginia' - Extracts from W.Bradford's 'Plymouth Plantation' - Extracts from Thomas Paine's Common Sense -Washington Irving's RipVan Winkle and Edgar Allan Poe's The Fall of The House of Usher

Table 2.5 - The programme of the second year level

In the third year, the learners are taught approximately the same modules as the second year. The module of written expression is included to improve the learners' writing level. Its content deals with the production of essays through the teaching of the different types of writing as well as the way of organizing compositions. The module of oral expression includes discussions, dialogues and role plays for the purpose of developing the students' speaking skill. The module of psychology is added at this level. Its content aims at preparing the learners for their future career as teachers.

Third world literature is introduced at this level to enable the students to study the African oral tradition and analyse some poems written by African poets.

The programme of the third year is provided in the table below.

Modules	Lectures
Linguistics	- Linguistics and sociolinguistics - Dialectology - Anthropology - Sociology- Language variability - Language versus dialect - Dialect and accent
British Civilization	- Trade unionism - The Grand National Consolidated Trade Union -Chartism - The cooperative movement - Foreign trade and free trade
American Civilization	- The articles of confederation - The different administrations of American presidents - Sectionalism - The civil war - The reconstruction
British Literature	- 18th century novelists (Bunyan, D. Defoe)- Defoe's <u>Robinson Crusoe</u> - Romanticism - First generation of romantic poets (W. Blake , W. Wordsworth , S.T. Coleridge) - Second generation of romantic poets (Lord Byron , Shelly , Keats) - Early Victorian writers - Jane Austen - Austen's <u>Mansfield Park</u> - Victorianism - Characteristics of Victorianism - Victorian Writers - Victorian poets (Mathew Arnold, Thomas Hardy, Browning, Christina Rossetti) - Victorian novelists - Charles Dickens's <u>Bleak House</u> - George Eliot's <u>Silas Marner</u> - Criticism in literature

American Literature	- Realism - The genteel tradition (Henry James's Daisy miller) - The local colour movement (Mark Twain's The Adventures of Huckleberry Finn) -The Afro -American literature - Slave narratives (F. Douglas's Narrative of the life of Frederick Douglas, An American Slave) - Harlem Renaissance - Protest literature (R. Wright's The Ethics of Living Jim Crow) - Integrationist literature (J. Baldwin's The fire Next time)
Psychology	- Humanistic education - Negotiation and Responsibility - Facilitation and management of learning - Teaching and learning - How are languages learned? - Nature of input - Process of intake - Interaction in the classroom - The role of errors - Effects of learners' differences on learning - Learning styles and strategies - Affective factors -Motivation - Teacher's roles and responsibility- Learner's autonomy and training - Roles of learning materials

Table 2.6 – The programme of the third year level

The programme of the fourth year gives the learners broader information related to the modules instructed previously with the exception of the four skills that are not designed to be taught as separate modules despite their occurrence in the teaching of the other content areas. The students are also presented to new concepts covering the field of third world literature and civilization.

The programme of this level is elicited in the following table.

Modules	Lectures
linguistics	- Brain and language - The modularity of the Brain -The autonomy of language - The evolution of language - Language acquisition - Theories of child language acquisition
Literature	- Modernism - James Joyce's The Dead (from <u>Dubliners</u>) - J. Conrad's <u>Heart of Darkness</u> - The 20th century poetry -Poetry of the first world war - Poetry of the 1920's (TS.Eliot and J.B Yeats) - Poetry of the 1930's (The Auden generation)
Third world literature	- The negritude movement - Study and analysis of Chinua Achebe's <u>Things Fall Apart</u> - The study of short stories
African Civilization	- The partition of Africa - The Berlin conference - Africa during the two world wars - Nationalism in Africa – Panafricanism
TEFL	- Teaching English - Language and teaching - The teaching and learning process - Language teaching methodology - Learning styles and learning strategies
Psychology	- Definitions of language - Language acquisition - Linguistic competence and performance - Foreign language teaching approaches - Interactions - Language and culture -Taxonomy of foreign language strategies - Types of foreign language learning strategies.

Table 2.7- The programme of the fourth year level

During the four years, the learners are supplied with a wide variety of information about the English language. The purpose of the inclusion of the designed modules is to make the students master the four skills and get sufficient content knowledge concerning specific areas or disciplines within the target language.

2.9- The Evaluation of the Students

The students are evaluated for placement or certification purposes. Thus, the process of evaluation at university aims at placing the learners in the following levels as far as the first, second and third year students are concerned in addition to awarding a licence degree to the fourth year learners. For each level, the same steps are involved. The learners' evaluation comprises three phases. The first phase includes two yearly examinations.

The second one involves a comprehensive examination assessing only the students who failed in the two former examinations. The third phase is a make-up examination intended just for those learners who failed in the previous examinations.

In addition to the official examinations whether oral or written, students' projects are counted as a form of evaluation. Also, the fourth year extended essays or training reports are scored and used as a form of assessment that is complementary to the results obtained in the examinations related to the other modules.

The principles of the ministry of higher education concerning the evaluation of the students state that:

> *The continuous evaluation of the students' aptitudes and knowledge can*
> *be carried out by:*
> - *Written examinations of average duration ranging from one to two*
>   *hours.*
> - *Short-term projects*
> - *Long-term projects*
> - *Evaluations of practical work*
> - *Evaluations of guided work or any kind of participation.*
> - *Memoirs or training reports evaluated separately.*

Therefore, there is one examination taking place in each semester which leads the students to get two marks at the end of the year. The mean of these scores represents the final grade.

According to Article n°8 of the national charter:

> *The final mark of the module is obtained through the evaluation of written examinations and projects (T.D).*
>
> *The general mark is counted by combining the grades of all modules obtained after calculating the mark with its coefficient concerning each module.*

After obtaining the results of the two examinations by the end of the two semesters, another test called a comprehensive or qualifying examination takes place in June to test those students who did not get the average during the year. The final mark is calculated using the following method:

$$\frac{((\text{comprehensive examination} *2) + \text{project}) * \text{coefficient}}{3}$$

The grade obtained is compared with that of the year and the general mark determines if the student passes or fails. Article n°10(of the national charter) states that:

> *The admission to the superior year is pronounced when the student obtains:*
> - *general average equal or superior to ten out of twenty (10/20)*
> - *and an average at least equal to five out of twenty (05/20) in each module.*

Finally, a make up examination takes place in September to evaluate those students who failed in the previous examinations. Article n°10 (of the national charter) explains that:

The student who was not admitted in the following level, can sit for a second session of examinations organized in September... In this session, the student attends examinations of the modules in which his average is inferior to ten out of twenty (10/20).

This article adds that '*at the end of this make up session, the general average is calculated according to the best marks obtained in the examinations of September and those obtained in the examinations of June.*'

The evaluation of the students at university is totally different from the other levels of education that are considered as a means of making learners reach the field of higher education. At university, the student has more than one opportunity to make a remedy to his failure in examinations. These opportunities are embodied in the structure of a synthesis and make-up examination.

In addition to this, examinations are not the only means of evaluating the students since projects, personal presentations and dissertations or memoirs are also taken into consideration.

2.10- Testing English at University

The students of English as a foreign language are tested at university through two yearly examinations in addition to two additional make up examinations that are used to assess those learners who did not get the average during the year.

These examinations aim at testing the learners' ability to use the English language and their degree of mastery of what was taught to them in order to allow them to pass from a level to another. The main purpose of assessing the first, second

and third year is to decide about their placement in the following level. The purpose of testing fourth year students is to make a decision concerning those learners who will get a degree of English and those who will not. Although the number of examinations taking place at each level is the same, the types of questions concerning each module differ from a level to another in their form, content and degree of difficulty.

Examinations of English at the university level can be grouped into two categories which refer to the assessment of the oral and the written form of the target language.

2.10.1- Oral Examinations

The spoken language is assessed within the module of oral expression. In this kind of examinations, the main concern is to test the students' ability to speak English fluently taking into consideration their way of pronouncing words and the extent to which they are able to produce acceptable and meaningful sentences.

The questions involved in such examinations often require the students to narrate a story, to discuss and state one's opinion concerning a given topic , to explain certain proverbs or take part in a dialogue or a play.

Most of the time, the learners are requested to speak for a certain period of time ranging from five to fifteen minutes.

In addition to the oral expression examinations, it can be said that the students' ability to use spoken English is also assessed when the learners are asked to make oral presentations in which they summarize and explain the content of written projects dealing with specific subjects in certain modules such as civilization and literature.

Also, the fourth year students who produce extended essays are required to make up in an oral presentation during the viva voce so that a general mark can be awarded to the learners on the basis of the written form of the extended essay and the oral presentation.

2.10.2- Written Examinations

With the exception of the module of Oral expression, all the other modules are written. They aim at assessing the learners' ability to use the English language in a written form and their mastery of the content of the modules.

These examinations last for one hour and a half. They vary from a module to another in their form and content. The examinations concerning each module differ from a level to another not only in their content and the type of questions used but also in their degree of difficulty.

Testing the first year students involves the use of variety of exercises that require separate answers in the examinations of the majority of the modules. However, the learners need to write an essay of a considerable length in the examination of linguistics.

As far as the second year is concerned, one can say that the majority of the examinations compel the learners to write an essay while the modules of grammar, written expression and phonetics include exercises needing short answers. In the third year, only the examination of phonetics is composed of separate exercises while all the other examinations involve questions urging the students to write essays. The fourth year examinations are in the form of essay questions.

2.11- The Different Types of Questions Involved in Each Module

The examination related to each module is different from those of the other ones not only in content but also in the type of questions employed.

The questions involved in the oral expression examination mainly focus on making the students discuss topics, tell stories or act out a dialogue.

On the contrary of the other examinations, this type requires the student to answer only one question within a limited period of time. Since the learners do not sit for the examination at the same time, the form of the question is kept but the content varies from a student to another. For example, if one learner is urged to discuss a

certain topic, most of the other students will be asked the same question. However, the topic dealt with will differ from a learner to another.

The reading comprehension module is assessed through the use of short texts accompanied by several exercises inviting the students to answer questions related to the text, to provide the meaning of some underlined words according to the given context, to choose the correct answer and finally to write a short paragraph in which the students are either requested to summarize the text or to give their point of view concerning the discussed topic.

The examination of phonetics is based on the use of two types of exercises. The first type includes questions related to the theoretical part of the lectures in which the students have to provide definitions concerning the concepts studied, to describe speech organs, to explain the role of these organs, to describe consonants and vowels and to explain the difference between them. The second type of exercises reommends the students to transcribe vowels, consonants or words.

The module of grammar is tested through the use of separate activities involving short answers. In the examination of this module, the students are given exercises that urge them to find sentence patterns, write the correct tense of the verb and to transform sentences from passive to active or vice versa. They may be also asked to transform sentences from the direct to indirect speech, to fill in the gaps with articles or prepositions, to derive nouns from verbs, to derive adjectives from verbs or nouns, to find the types of sentences and to ask questions. Not all these activities are involved in the same examination as only three or four types of these questions are used so that to avoid the problem of time constraints.

The examination of written expression requires the students to supply punctuation and capitalization, to find the type and function of subordinate clauses, to correct mistakes, to provide definitions of words and to write sentences using the given words. The examination includes two or three options from these exercises. In

addition to this, the students are asked to write either a paragraph or a dialogue involving a specific topic to be discussed.

The modules of linguistics, civilization, literature, TEFL and psychology urge the students to write an essay. However, the type of the essay question varies in its form from a module to another despite the fact that the purpose of the use of an essay question for these modules is to make the learners summarize the lectures studied in each specific module.

The examination of American civilization relies on the employment of direct questions introduced by the words "what, why, how". In British civilization the question is in the form of a quotation followed by the word 'Comment'. This method urges the learners to summarize the lecture and synthesize it at the same time.

The module of British literature is assessed by asking the students either to study and analyse a poem or to answer direct questions introduced by "how, why, what, comment, do you think…?". In both cases, an essay must be provided involving not only the knowledge learnt from the lectures but also the learners' personal opinion about the topic.

The examination of American literature makes the students write an essay after choosing one of two questions. The first question involves a quotation followed by the word 'Comment'. The second quotation is in the form of a short passage taken from the studied texts or short stories followed by a set of questions concerning the author, the theme, the style and the literary devices used in this passage. The answers of these questions should be provided in an essay.

The examination of linguistics takes a different form according to the tested level. The first year students are assessed through the use of an exercise including a direct question or a quotation followed by the word 'comment' to be discussed in an essay in addition to a second exercise which requires the students to provide

CHAPTER THREE:

CASE STUDY:
DATA COLLECTION

CHAPTER THREE :
CASE STUDY: DATA COLLECTION

3.1- Introduction

This chapter is devoted to the description of a case study concerning the use of holistic scoring for the assessment of students' writing performance taking as a sample the fourth year students from the English section at the University of Tlemcen. First, it explains the objectives of this research work and describes the research design. It also speaks about the sample population and the variables involved in this case study. Finally, the main research instruments used to achieve this work are described in this chapter.

3.2- The Research Objectives

This research work deals with the case study of the assessment of fourth year students at Algerian universities. It is related to several objectives.

First, this project aims at providing data about the practical field of the assessment based on holistic rating. It tries to find out the main methods used for correcting essay examinations focusing mainly on the way of rating content and form. This research work is meant to constitute a source of information concerning the students' difficulties and ways of perceiving essay examinations. It also tries to provide a broader view concerning the learners' errors and the teachers' way of dealing with holistic rating taking into consideration how they are assessing grammatical competence as well as the mastery of the taught modules.

In addition to this, this project aims at supplying the reader with certain details about the problems encountered by teachers when correcting essay examinations. The gathered data can help to provide some suggestions that may alleviate such difficulties and decrease teachers' subjectivity.

The main purpose of this research work is to give a description of the assessment based on holistic rating concerning EFL learners with special reference to the case of fourth year students of English at the University of Tlemcen studying mainly the difficulties of both the learners and teachers when dealing with essay

examinations. It attempts to shed light on the extent to which these learners are able to perform well in such examinations.

3.3- The Research Design

Research means the act of investigating about a certain phenomenon using observation, experimentation and logical analysis to get sufficient information about the problematic issue in order to draw conclusions and try to provide suggestions for the improvement of the existing situation.

Educational research has often been characterized by two different paradigms which are positivism and interpretivism. Positivism means a scientific approach. It stems from the idea that true knowledge is objective. The positivist researcher looks for quantitative data and generalization. Science and scientific studies are positivist as they are objective and separate facts from values. On the other hand, interpretivism relies on the exploration of perspectives to develop insights into specific situations. It is based on qualitative data and interviews. The case study is incorporated in the interpretive paradigm as it acknowledges the researcher's subjectivity and searches for personal knowledge (Wellington, 2000).

The idea of the interpretive versus positivist view is closely linked to the existence of qualitative and quantitative research approaches. Qualitative research involves a variety of empirical materials such as observation, personal experience and interviews. It is associated with interpretivism which perceives reality as a construct which people interpret in different ways. It is employed in the field of education mainly because the educational researcher needs to pursue research that is grounded on people's experience.

Quantitative research refers to those approaches that are based on the assumptions of scientific investigation. It puts the emphasis on patterns, causes and consequences by applying the rules of positivism. It relies on data collected from tests, experiments, questionnaires and interviews (Scott and Morrison, 2006).

Qualitative methods of data collection search for image-based evidence while quantitative methods of data collection focus on numerical evidence. However, qualitative methods should not be totally value-laden or subjective as they must rely on a certain theory and hypothesis. In the same way, quantitative methods can never be totally value-free or objective since they are employed by human beings. This means that these two approaches have to complement each other to produce successful research (Wellington, 2000).

In the domain of education, it is assumed that quantitative and qualitative methods are totally different as they approach educational phenomena in distinct ways. Therefore, researchers propose the employment of mixed methods in educational research. Mixed methods refer to the combination of qualitative and quantitative methods in order to obtain better results.

The process of data collection involves the use of qualitative and quantitative information. Also, qualitative instruments such as interviews can be utilized to collect quantitative data and quantitative tools like questionnaires can provide qualitative information. In fact, *"...all the methods of educational research deal with qualities, even when the observed qualities are counted. Similarly, all methods of analysis use some form of number"* (Gorard, 2001: 6). In qualitative analysis these forms of number may be formulated through the words 'tend, most, some, all, none, few'. They constitute the patterns of qualitative analysis. Therefore, qualities are numbered to provide an indepth description of the studied element. This means that qualitative methods can use numerical data. It should be mentioned that qualitative analysis is characterized by specificity but quantitative analysis can provide a basis for making generalizations.

The effective use of qualitative and quantitative data is closely related to the validity of research. In qualitative data, validity is related to richness, honesty and depth of the gathered information as well as the objectivity of the researcher. The validity of quantitative research is based on the sampling strategies and the statistical procedures employed for data analysis. However, it is believed that research cannot be totally valid because of the sampling errors and the subjectivity of the

respondents whose opinions and attitudes influence the research results and produce bias. Thus, validity is not an absolute state but it can be maximized by relying on more quantitative methods and trying to minimize bias (Cohen et al, 2007).

The choice of the research approach employed depends on the nature of the element studied. Thomas (1998) identifies six types of research approaches which are often referred to as approaches to data collection. These research approaches are: historical analyses, ethnographies, surveys, correlational comparisons, experiments and case studies. A historical analysis constitutes an investigation about specific cases located within a particular era of the past. It takes the form of an account concerning individual persons, institutions, educational practices or events. Ethnography is an approach used by anthropologists. It is a descriptive study of a particular society. The purpose of ethnographic research is to depict the culture of a group of people.

A survey provides a description of the characteristics of a group of persons, policies, practices or institutions. In educational research, a survey gives an overview of the status of some aspects of education. A correlational comparison focuses on gathering information about two entities and comparing them to elicit if the change occurring in one set has affected the other set. An experiment aims at discovering how two or more individuals or groups are influenced by a specific treatment.

A case study involves the analysis of a single entity which may be a person, a group, a system or a method by making description, explanation, evaluation or prediction in order to draw conclusions concerning a broader population affected by the studied subject. The case study is considered as a type of interpretive research involving a detailed examination of an entity through various research tools like observation and interviews. It plays an important role in educational and psychological research. This approach is not characterized by its methodology but by the object of study taking the form of a complex instance that has to be understood via extensive descriptions and analysis of this case or instance.

The case study is often viewed as a strategy that is employed in qualitative research. When studying a case, the collected data should provide information about the nature of the studied object, its background and those informants who enable the researcher to know about the case. The design of a case study has to take into consideration three main elements: the elicitation of the research questions, the identification of the hypotheses and the specification of the unit of analysis (Mertens, 2005). The research questions guide and direct the researcher in his work. Questions introduced by 'how' and 'why' are seen as the most appropriate for case studies. The research hypotheses explain why the researcher believes that he might observe or study a specific instance. They give the possibility to narrow the focus of the study.

Finally, the identification of the unit of analysis such as a group of individuals, a program or a course leads to the specification of the case and the establishment of clear boundaries for the research work.

Among the approaches discussed above, the case study is frequently used in educational research. It is very appropriate for such a situation as it provides an overall holistic presentation and assessment of the studied subject (Stufflebeam, 2000). Cohen et al. (2007) assert that the case study is a specific instance designed to illustrate a more general principle. This approach enables the researcher to draw conclusions and make generalizations from the analysis of a particular phenomenon or situation.

Therefore, the approach used to gather data for the present research work is a case study since it investigates about a single entity represented by the fourth year university students. It is employed for descriptive, explanatory and evaluative purposes. The research work features description by providing an account of the literature review concerning assessment as well as an overview of the testing situation at the Algerian university. It gives explanations of the research questions and hypotheses by eliciting the cause/effect relationship between the different variables and identifying the real state of the target variable represented by holistic assessment.

Finally, a sort of general evaluation is made about the efficiency of using holistic scoring for assessing students' essays by displaying the strengths and weaknesses of such a procedure.

This case study is composed of two main phases. The first phase is data collection which refers to the act of gathering information about the studied domain. The collected information can be qualitative or quantitative. Qualitative data give a general view or opinion about the individual or element studied. For example, saying that a learner is a good writer or not provides qualitative information. Quantitative data are in the form of statistics as they provide numbers or percentages concerning the investigated topic. For instance, quantitative information may include the scores of learners or the number of students performing well.

Both qualitative and quantitative information are furnished. Data are collected through the use of various research instruments including interviews, questionnaires, tests and document analysis.

The second phase of the case study is the data analysis and interpretation. At this level, the gathered information are analysed by making calculations and using statistics to obtain the final results of research. Thus, this phase involves employing statistical methods which are classified into two groups: descriptive and inferential. Descriptive statistics summarize information in a quantitative form while inferential statistics provide an estimate of how a sample accurately represents a broader population (Thomas, 1998)

The analysis and the interpretation of the data collected for the present case study rely on the employment of descriptive statistics which include frequencies, percentages, tabulation, graphic representation and measures of central tendency.

Frequencies represent discrete or enumerative data. They involve a counting operation providing a number about how many individuals or elements fall into a certain category or classification. *"The numbers of discrete units in a given classification are referred to as frequencies"* (Wilson et al., 1969: 502). They

indicate the relative frequency of the various variables to one another. Frequencies are also called ratios.

The numbers obtained from the data gathered can be transformed into percentages which tell the proportion (on a 100- unit scale) of a given variable (Thomas, 1998). A percentage is calculated by multiplying a ratio with 100.

The calculation of data enables to present it in the form of tabulation as the numbers obtained are classified in tables representing frequency distribution which refers to the number of times a certain element is repeated. Such tables are often called frequency tables.

Graphic representation refers to the demonstration of the frequency distribution in diagrams. Thus, the meaning of percentages is graphically conveyed so that data are immediately observed and understood. This method may involve the use of a bar graph which is in the form of a vertical line at the left side of the figure and a horizontal line at the bottom consisting of a series of bars representing statistical data. The bar graph is also called a histogram. It may be replaced by a frequency polygon which is a broken line graph connecting midpoints at a height that corresponds to the frequency. Graphic representation can be expressed through a pie chart which is a circular graph divided into sectors that show the proportional sizes of the quantities represented. A pie chart is also called a pie graph or a circle graph.

For the interpretation of the data gathered from the tests used as research instruments, specific formulas are utilized to determine the central tendency which refers to the average performance of a group of learners on the basis of the obtained scores. These techniques are called measures of central tendency. They are the mean, the median and the mode.

the total number of other persons or items (Thomas,1998). Thus, the median is the halfway score.

The mode is the number that occurs most often in a distribution of scores. For instance, in the set of scores: 12, 10, 9, 8, 6, 8, 7, 8, the mode is 8 because it is the most frequent score among all the numbers. The mode is the score obtained by the greatest number of students (Wilson et al., 1969; Thomas, 1998).

In addition to the measures of central tendency, measures of variability are also employed to determine the spread of the scores.

When analysing assessment results, it is important to take into consideration the variability of scores to understand the differences between the learners' marks. Variability, which is also called dispersion or spread, refers to the extent to which the scores vary from each other.

Although various measures of dispersion exist, range and standard deviation are two of the most common methods employed for indicating the spread of scores (Porte, 2002).

Range is a measure of dispersion that shows the variability between the highest and lowest scores. It enables the researcher to get an idea about the degree of differences in the learners' marks. *"Range is calculated by simply substracting the lowest score from the highest"* (Whiston, 2008: 3). Thus, it gives information about the width of variation between the scores.

On the other hand, the standard deviation demonstrates how the test results vary from the mean. It is often linked to the variance of scores which is the square of the standard deviation (Cumming and Berwick, 1996).

The variance is also called the mean square deviation. It is calculated as follows:

$$V = \frac{\sum(X - M)^2}{N} \quad \text{(Whiston, 2008)}$$

X = the score

M = the mean

N = the total number of scores

V = the variance

The standard deviation is the square root of variance. It is represented as:

$$S = \sqrt{\frac{\sum(X - M)^2}{N}} \quad \text{(Whiston, 2008)}$$

When the standard deviation is large, the scores are more deviated from the mean. If it is small, the marks are very close to the mean. The standard deviation provides data about the variation in the scores.

Generally speaking, this case study unveils a mixed-method approach that provides both qualitative and quantitative data.

The pivoting source of the process of data collection is the sample population from which information are elicited to analyse the case of the holistic assessment of writing.

3.4- The Sample Population

The sample population is a very essential element in the process of data collection. The term population is used to refer to the group about which the researcher has to make generalizations. *"Population is the theoretically specified aggregation of study elements... A study population is that aggregation of elements from which the sample is actually selected"* (Babbie, 2009: 199).

Thus, people or certain categories of people are the elements of study that constitute a target population which denotes a specific collection of persons that are concerned

with the findings of the undertaken research. The target population is often very large which makes it inaccessible. Within this population, the researcher identifies a smaller group of people referred to as the accessible population or the study population which delineates certain specifications of the target of research. Finally, a smaller group called the sample is chosen from this category to limit the field of study. The sample is a representation of the accessible population as it reflects the features of the population (Houser, 1998).

Researchers conduct research about the abilities, attitudes or behaviour of the target population. *"They often need to be able to obtain data from a smaller group or subset of the total population in such a way that the knowledge gained is representative of the total population under study. This smaller group or subset is the sample"* (Cohen et al., 2007: 92). Therefore, the sample is the source of information required to answer research questions and test the research hypotheses (Perry, 2005).

The main characteristics of an appropriate and useful sample are: representativeness and generalizability.

Representativeness means that the sample includes the same distribution of characteristics as the total population. It is achieved via probability sampling. However, this quality is not available in all respects as it is limited to those features that are relevant to the study. Thus, *"...a sample is representative of the population from which it is selected if the aggregate characteristics of the sample closely approximate those same aggregate characteristics in the population"* (Babbie, 2009: 198).

Representativeness is often implied from the analysis of the sample which leads to make assumptions about the degree of similarity existing in the features of the selected groups and the total number of the people who are concerned with the generalizations that have to be made from the study. It is a very essential element

that should be taken into consideration when selecting a sample which plays an important role in the process of data collection and analysis.

The employment of procedures of selection that focus on decreasing bias and representing the larger population is very necessary as the issue of representativeness provides a basis for the generalization drawn from the research results.

The sample should possess an important characteristic which is referred to as generalizability. *"Sample generalizability is the ability to generalize from the sample to the larger population and rests on the basis that the selected sample is representative of the larger population"* (Scott and Morrison, 2006: 219). This means that the sample has to enable the researcher to formulate descriptions that portray the main characteristics of the population. Therefore, the data gathered have to provide the possibility to make generalizations by asserting that certain statements referring to a group of individuals also fit to a wider population or category to which this group belongs.

In fact, the purpose of educational research is to make generalizations about the individuals or activities related to the context of education. Generalization can be based on two categories of inferences which include probabilistic and empirical generalizability. Probabilistic or deterministic generalizability involves quantitative attributes. It depends on the selection of a typical sample to ascertain that the case studied will be representative of the larger population that is concerned by the intended generalizations. Nevertheless, it is claimed that it is impossible to produce exact extrapolation from the sample to the whole population. Thus, probabilistic generalizability is based on inferences that represent near approximations reached via statistical analyses (Eisenhart, 2008).

Emperical generalizability relies on the examination of qualitative attributes. Thus, it operates at the level of experiences and is not concerned with ontology which means the claims or assumptions made about the nature of social reality through the utilization of a specific approach to social inquiry (Crotty, 1998). Empirical generalizability does not put the emphasis on theoretical perspectives as it depends on the observation and analysis of particular cases for the purpose of extrapolating the results from the studied instances to the population. However, *"...the relation between the sample and the population is bound to be approximate"* (Scott and Morrison, 2006: 118). This is why the researchers propose the utilization of the term transferability as an alternative to the notion of generalizability.

Transferability is linked to the examination of the extent of similarity between two contexts. It refers to the idea that the results obtained from one context can be transferred to another one if these two contexts are sufficiently identical. It does not search for abstract universals achieved by statistical generalizations, but it looks for concrete universals reached by studying a particular case and comparing it with other studied cases (Eisenhart, 2008). Transferability means that the task of finding the proof for the significance of the research results is left to the readers or users of the research project. As a result, the researcher has to provide a profound description of the research problematics and findings in order to enable the readers to establish a relationship between the sample and the population (Scott and Morrison, 2006).

Another issue that is related to sampling is the homogeneity of a sample. A homogeneous sample includes members or elements that are alike on one dimension or more. In reality, *"...a homogenous population produces samples with smaller sampling errors than a heterogeneous population does"* (Babbie, 2009: 215). This means that the homogeneous sample allows the researcher to draw generalizations that can approximately reflect the characteristics of the total population.

The fact that the elements involved in the study possess similar characteristics facilitates the researcher's task and enables him to draw certain conclusions that can be accurate to a certain extent since the degree of homogeneity can lead to control the variable under investigation.

The major decision that is closely related to the issue of sampling is the sample size which remains a question that does not have a definite response. *"There is no clear-cut answer, for the correct sample size depends on the purpose of the study and the nature of the population under scrutiny"* (Cohen et al., 2007: 101). Even if there is not a total agreement concerning this matter, the size of the sample has a great effect on the research results as it determines the extent to which this sample is representative to the total population as well as the accuracy of the statistical inferences. As a result, whenever the sample is larger, the sampling error is smaller and there is a greater chance of obtaining significant results (Cottrell and McKenzie, 2005).

However, the sample may be composed of a minimum number of cases if the researcher intends to employ the data for statistical analysis or to examine relationship between subgroups. Therefore, the researcher has to take into consideration the variables to be studied, the statistical procedures employed for data analysis and the subgroups of the population in order to select a sample that suits the research context. In fact, a sample that is too large might be unwieldy and a sample that is too small might be unrepresentative (Cohen et al., 2007)

This means that a practical study is based on a sample that is large enough so that it is representative of the population and gives the possibility of making generalization, but this sample should also be small enough in order to save time and facilitate data analysis.

Nevertheless, the size of the sample is determined by the nature of the undertaken research. Thus, certain cases may need small samples while others may require larger ones. For example, a small sample is needed when studies are pilot, when the research is exploratory, when the data collection techniques are time

consuming, when statistical procedures are utilized or when resources are limited. On the other hand, a large sample is required when employing surveys, when expecting high attrition that may be in the form of non respondents, when a large number of extraneous variables exists, when the units of the sample display heterogeneous variables or when the number of the analysed variables is very large (Gorard, 2001; Cottrell and McKenzie, 2005; Cohen et al., 2007).

Therefore, the researcher does not need to focus too much on the size of the sample. *"The quality of a sample bears little relationship to its size proportionate to the population. A sample does not have to be a certain proportion of the population"* (Gorard, 2001: 18).

The researcher has to design a sample that is suitable to the research objectives and methods. The determination of the size of a sample may vary according to the type of the needed information. The provision of qualitative data may require a different sample size than the one which is employed for obtaining quantitative data. In a qualitative study, no definite sample size is distinguished. For instance, a sample of five individuals out of a population of thirty persons is seen as sufficient (Cohen et.al, 2007).

The burden of making a decision about the number of cases that should be included in the sample is left to the researcher.

> *There are no rules for sample size in qualitative inquiry. Sample size depends on what you want to know, the purpose of the inquiry, what's at stake, what will be useful, what will have credibility, and what can be done with available time and resources... In depth information from a small number of people can be very valuable, especially if the cases are information rich.* (Patton, 2002: 244)

The results obtained from qualitative data do not rely on the sample size but they depend on the quality of the provided information and the procedures that are employed by the researcher in order to analyse the gathered data. Thus, a small or large number of cases can be involved in the study. Nevertheless, the issue of the homogeneity of the sample can be a major factor in making decisions concerning the sample size. This means that a small sample is selected if there is homogeneity in the population while a large sample is opted for when there is heterogeneity in the population (Cohen et al., 2007).

In quantitative studies, three theoretical considerations are viewed as a basis for determining the size of the sample. These considerations are: the central limit theorem, precision and statistical power analysis.

The central limit theorem (CLT) is related to the fact that the mean of the sample should approximate the mean of the population for a specific feature. Thus, the characteristics of the sample have to portray the general features of the population. This can be achieved whenever the size of the sample increases. Precision refers to the use of statistics to calculate the sampling error (Cottrell and McKenzie, 2005).

Sampling error means that the process of generalizing from a sample to a population often involves a certain error. There are two kinds of error: sampling and non sampling error. The sampling error which is also called random error may result from the fact that the sample is composed of too many elements of one kind and less elements of another kind. Sampling error decreases when the sample size increases. It is also small if the population is homogeneous (Dattalo, 2008). Sampling error means the difference between the mean of the sample and the mean of the population. It exists because of the differences existing between the informants. Thus, it depends on the variability or heterogeneity in the population.

In order to calculate how accurate the results of the study are, the standard error of proportions is employed to find out if the proportions of responses are truly representative of the total population. The standard error of proportion is calculated as follows:

$$SE = \frac{\sqrt{p \times Q}}{N}$$ (Cohen et al., 2007)

P = the percentage in favour of the stated issue or topic,

Q = 100 percent - P,

N = the sample size.

SE = standard error

The standard error of proportion is one of the various techniques used to identify sampling error. The degree of sampling error depends on the homogeneity of the sample as it is linked to the extent to which the characteristics of the members of this sample are identical. The increase of sample size leads to the decrease in sampling error.

On the other hand, non sampling error, which is also called systematic error, constitutes a serious problem to the researchers as it cannot be controlled by increasing the size of the sample. This kind of error is referred to as bias. There are three types of non sampling error which include selection bias, non-response bias and response bias. Selection bias refers to the tendency to exclude certain elements from the sample. It can be minimized via probability sampling. Non response bias means that non-respondents may differ from the informants. As a result, the extrapolation from respondents to non-respondents becomes a problematic task. Response bias refers to the idea that the wording of the questions or the attitudes of the investigator may lead the informants to provide certain answers that can deviate from reality (Dattallo, 2008).

The sampling error can affect the accuracy of the research results. It is inferred from the characteristics of the sample and those of the population. It is estimated via inferential statistics. The degree of precision of sampling is based on the estimation of the sampling error which evidently decreases when the sample size is large.

The other element that is important in determining the size of the sample is statistical power analysis which is linked to the use of tests as research instruments. The power of a test refers to the degree to which it is able to provide accurate information. Statistical power analysis is *"... a measure of the test's ability to separate out genuine effects from random variation"* (Gorard, 2001: 19). This means that a powerful test can be a valuable research tool as it provides data that contribute to the explanation and elicitation of the problematics. Statistical power analysis is based on four factors which are: the sample size, the level of significance, directionality and effect size. The sample size is associated with statistical power analysis as the power of the test increases when the size of the sample increases.

The level of significance refers to the probability of making an error in statistical calculations. Directionality refers of the researcher's prediction about the direction of the outcomes of the employed instrument. Effect size is related to the researcher's expectations of the strength of the outcomes of the study which has to be based on the results of previous research (Houser, 1998). Therefore, central limit theorem, precision and power analysis are three basic considerations involved in the process of sampling. The researcher has to take into account one of those elements in order to determine the size of the sample which is the basis of research as the reliability of the statistical inferences is associated with the degree to which this sample is representative of the population (Cottrel and Mckenzie, 2005).

There is no evidence that a specific size is totally representative and yielding to generalization. However, it is assumed that the confidence interval can be used as basis for estimating the sample size. The confidence interval is referred to as the standard error which is equal to the standard deviation. It is a measure of variability that is calculated via a specific formula (see page 104). The confidence interval indicates the accuracy of results so that to enable the researcher to draw generalizations from the sample to the population.

Even if there is not a specific proportion that denotes the sample size, Gorard (2001) suggests that a good sample is the one that is five or ten times the number of variables employed. Mertens (2005) advocates the idea of choosing the smallest sample that takes into consideration the variability of statistical measures as well as the issue of sensitivity which refers to the probability of finding a difference when it exists. Dattalo (2008) assumes that the results are more likely to represent the population when the sample is large. However, *"...a large sample is no guarantee of accuracy"* (Balnaves and Caputi, 2001: 93). There are certain factors which prevent the obtention of a large sample. Attrition is an example of these factors. Thus, any sample size that is selected represents the maximum that a researcher can achieve (Gorard, 2001).

The purpose of sampling is to select a subset of the population to produce a practical study that allows generalization from the sample to the population. The researcher has to design the size of the sample that will lead him to achieve this purpose. Nevertheless, this is not an easy task. *"The issue of sample size is a lot like the problem students have when they are assigned an essay to write"* (Patton, 2002 : 245). This means that the decision related to the number of persons included in the population has to take into account the context of the study, the nature of the informants and the techniques used for data collection. This decision should not involve samples that are irrelevant (including two or three individuals) or time consuming (including a very large number or the whole population).

There are various types of samples due to the distinction in the sampling methods and the sampling designs. The ultimate purpose of sampling is *"...to select a set of elements from a population in such a way that descriptions of those elements accurately portray the total population from which the elements are selected"* (Babbie, 2009: 199).

There are two kinds of sampling methods: probability and non-probability sampling. Non-probability sampling refers to any sampling method relying on the selection of

elements possessing certain criteria. Thus, some units in the population have no chance of being chosen. In this case, the selection is non random. Non probability sampling is often used in situations where the sample is selected without the use of probability theory. There are four types of non probability sampling including convenience sampling, purposive sampling, snowball sampling and quota sampling (Babbie, 2009).

Convenience sampling, which is also called accidental or haphazard sampling, means that the researcher gets information from those people he encounters in society. Therefore, the sample is built up and does not represent the whole community. It only represents itself (Bouhouche, 2002). Convenience sampling relies on available subjects. The sample is often obtained by stopping persons in the street or other places to ask them questions.

Purposive or judgemental sampling is based on the selection of a sample according to the objectives of research and the knowledge of the elements of the population (Babbie, 2009). It involves studying cases on the basis of their typicality to form a sample chosen for a specific purpose. It illustrates a feature or process in which the researcher is interested (Silverman, 2000). The researcher tries to make an investigation which involves a phenomenon that is most likely to happen which means that the sample is carefully selected to achieve the research purposes. Purposive sampling is very useful for the researchers who want to get information about a small subset of a larger population or to study deviant cases that are not in harmony with regular patterns of attitudes.

Snowball sampling refers to the identification of a small number of people who have the characteristics in which the researcher is interested. *"This procedure is appropriate when the members of a special population are difficult to locate such as homeless individuals, migrant workers or undocumented immigrants"* (Babbie, 2009: 193). These people are employed as key informants to put the researcher in touch with other persons who will identify others. This method is used when access

to a population is difficult because the topic, for example, is sensitive (Cohen et al., 2007).

The term snowball is utilized to refer to the idea that the respondents suggest other informants leading to the process of accumulation of information via a few members of the target population.

Quota sampling means dividing the population into non-overlapping subgroups from which elements are selected on the basis of a specified proportion. It is widely used in market research (Sapsford, 2006). Quota sampling relies on a table which describes the features of the target population and establishes the proportions that fall into various age categories, educational levels or ethnic groups (Babbie, 2009). Therefore, all these types of non-probability sampling are related to specific criteria.

On the other hand, probability sampling, which is also called random sampling, involves random selection. It is the primary method of selecting representative samples as it increases the opportunity of giving an exact description of the population via the features of the chosen elements. Probability sampling is based on a probability theory which is a branch of mathematics that supplies the researchers with sampling techniques that can be utilized for the design of samples. *"More formally, probability theory provides the basis for estimating the parameters of a population. A parameter is the summary description of a given variable in a population"* (Babbie, 2009: 200). Therefore, this theory allows probability sampling to produce samples from which generalizations about the total population can be drawn through statistical analysis. In this way, bias can be avoided to a certain extent.

The main principle of probability sampling is that representativeness can be achieved if all the members of the target population have the same opportunity of selection. This is why this procedure is referred to as equal probability of selection

method (EPSEM). Probability sampling relies on random selection which is employed for the purpose of decreasing bias on the part of the researchers. Random selection means a procedure where every member of the target population has a chance of being selected. According to Donnelly (2004: 158), there are four types of probability sampling. They are: simple random, systematic, cluster and stratified sampling. These categories refer to the different ways of gathering a random sample.

In simple random sampling, every member of the population has an equal chance of being selected as the sample is not specified in advance. This method may involve the use of a random number table to pick samples at random. Simple random sampling relies on assigning numbers to the units composing a population. Next, a set of random numbers is generated. Then, the elements having those numbers are chosen for inclusion in the sample (Babbie, 2009).

Systematic sampling employs a technique that relies on selecting every K^{th} member of the population to be in the sample. If 'N' is the size of the population and 'n' is the size of the sample, $K = \dfrac{N}{n}$ (Donnelly, 2004). Thus, the value of 'K' depends on the size of the sample and the size of the population. 'K' is referred to as the sampling interval. The first unit in the sample is selected at random (Babbie, 2009).

Cluster sampling implies the division of the population into groups or clusters, from which a simple random sample is selected to form the final sample. Stratified sampling depends on the organization of the population into categories or strata so that the researcher can randomly select samples from each of these groups.

As far as this research work is concerned, the main sampling method used was the simple random sampling underlying an equal probability of selection so that each person in the target population had the possibility of being included in this random sample. This method was opted for to minimize bias. In addition to this,

systematic sampling was employed for the selection of the sample of the students'
scores.

This case study took as a sample population the fourth year students of the English
section at the university of Tlemcen.

The sample population involves only the students of the fourth year because
they are assessed via the exclusive use of essay examinations which are linked to
holistic rating. Also, the fourth year students are supposed to have mastered the
language to a certain extent that is better than the other levels which leads to an
efficient analysis of their way of dealing with examinations through the study of
their written production whereas the other levels may not perform in the same way
as they do not possess the same knowledge of vocabulary and grammar.

Finally, the choice of such a sample was meant to limit the sphere of the
population studied because of time constraints and difficulties of gathering
information. The total number of the target population was 153 students. This
population was somewhat homogeneous since the learners had approximately the
same age with the exception of a few cases that might display a slight difference in
their age. Generally, the age of the learners ranged from twenty-one to twenty-four
years old. Also, this category of learners can be considered as homogenous at the
level of gender since the total number of fourth year students included 133 girls and
only twenty (20) boys.

Another aspect of homogeneity was revealed in the learner's linguistic
background which was assumed to be almost alike as they have been exposed to the
same teaching materials during the years of their study at university.

As far as the students' performance was concerned, it can be stated that this
variable could not be totally controlled as individual differences would always exist.
Thus, the learners' can be grouped into three main categories. These categories
included students whose achievement was low, students performing well and other
learners whose performance was simply average. The fact that the amount of

heterogeneity was small within this population provided the possibility of selecting samples whose size was not very large since the extent of sampling error decreased in homogeneous samples.

Within this target population, various samples were selected to collect a wide range of information that could provide a better understanding of the studied case. However, the size of each sample differed from the other because of several factors such as time restrictions, attrition, the degree of manageability of data analysis and the use of tests as a research instrument.

The researcher not only chooses the sample to be studied but also designates the variables that have to be examined.

3.5- The Variables

When examining the learners' performance, labels such as variables and constructs are used.

> *A variable is essentially what we can observe or quantify of the human characteristics or abilities involved, whereas a construct is the actual characteristic or ability that it represents in human beings.* (Brown, 1988: 8)

A construct refers to the abstract ability of the individual while a variable refers to the aspect that can be measured or observed within this ability. For instance, the learners' ability to learn a language is a construct since it is unobservable. However, the learners' proficiency or degree of achievement is a variable because it is observable and measurable through the use of scores.

Therefore, a variable refers to human characteristics or abilities that may vary throughout time such as the learners' proficiency or motivation (Scott and Morrison, 2006). Various kinds of variables can be found. Among them, one can list the dependent, independent and intervening variables.

The dependent variable refers to the changeable element that is examined in relation to other elements. *" A dependent variable is observed to determine what effect , if any, the other types of variables may have on it"* (Brown, 1988: 10). For instance, the researcher may decide to investigate about the learners' proficiency in learning a given language. This proficiency is a dependent variable as it is linked to the abilities of the learners and the way this language is taught.

The independent variable means the studied element that may cause other variables to change. It is selected and manipulated by the researcher to determine whether it has any effect on the dependent variable. For example, when testing the learners' proficiency, the independent variable would be represented by the specific period of time during which they studied the language since this time duration has an effect on the degree of the learners' achievement.

The intervening variable is a label used by the researcher to describe the relationship between independent and dependent variables. For example, when examining the learners' proficiency of a language after learning it for a specific period of time, the intervening variable can be language learning or the exposure to language. Thus, the intervening variable is an abstract label applied to explain the process that links dependent and independent variables. In fact, it is a construct that is not observable.

In this research work, the dependent variable is the students' performance represented by the marks of the learners as well as their degree of success in writing an essay. The independent variable is embodied in the programme or the taught material representing the input. The intervening variable is the students' exposure to the written form of the target language.

In addition to the variables mentioned before, Ruth and Murphy (1988) refer to the hidden variable in writing research. They assert that Braddock, Lloyd-Jones and Schoer were the first writers about research in composition who mention the hidden variable which is often ignored in research stating that this type of variable is represented by the assignment variable which has four aspects including the topic, the mode of discourse, the time afforded for writing and the examination situation. One can say that the topic effects constitute the hidden variable. In this research work, the hidden variable is the impact of the essay question on the learners' performance represented by the degree of easiness or difficulty of the writing task.

3.6- The Research Instruments

To achieve this research work, the use of different tools for gathering information was necessary. The instruments involved are: interviews, questionnaires, tests and document analysis.

3.6.1- The Interviews

An interview is a conversation taking place between the researcher and the informant involving a number of oral questions.

> *Interviews enable participants -be they interviewers or interviewees- to discuss their interpretations of the world in which they live, and to express how they regard situations from their own points of view.*   (Cohen et al, 2007:267)

The interview may be employed with other research tools to collect a wide range of information about the field studied. *"Often interviews are used where other research instruments seem inappropriate"* (Wilkinson and Birmingham, 2003:43).

In fact, interviews help the researcher to gather additional data as the interviewer can interact with the respondents. They may be labelled differently according to the way of administering the questions.

Three models for interviews exist. They are the structured, the semi-structured and the unstructured interview (Corbetta and Patrick, 2003).
The structured interview requires all the respondents to answer the same questions which have the same wording.

The semi-structured interview involves an outline of the topics that will be discussed. In this case, the interviewer is not obliged to follow the same order and wording of the given questions. On the other hand, the unstructured interview depends on the respondent's collaboration since the interviewer has just to explain the nature of the issues discussed. The form and the content of such an interview are not predetermined.

Different strategies are employed to design and prepare the questions of the interview.

> *When planning an interview, researchers can profit from deciding which of several strategies best fits their needs. Four popular strategies bear the labels loose-question, tight-question, converging-question and response-guided.*
>
> (Thomas, 2003:63)

The loose-question strategy, also referred to as broad-question strategy, aims at eliciting the informants' interpretation of general query. Thus, the questions involved in this approach ask the respondents to tell what a word or phrase means to them. A loose strategy includes open-ended questions that are in the form of ' Wh ' questions which require long answers.

The tight-question strategy employs questions that urge the informants to select answers from a set of given responses in the form of yes/no, like/dislike or other multiple choice options. This type of questions is sometimes followed by a request to explain why the respondent opted for that choice over the others. The purpose of the tight strategy is to discover the interviewees' preferences among a limited number of options.

The converging-question strategy which is also called the funnel strategy is a combination of loose and tight approaches. Thus, the interviewer uses open-ended and multiple-choice questions. The response-guided strategy includes the use of a prepared question followed by queries which are logical extensions of the respondent's reply to the initial question. In this way, the interviewee's answers provide opinions about some issues related to the opening question.

Regarding this research work, the interviews were used only with the teachers. Their purpose was to collect data that could not be provided by the questionnaire because of the constraints of the latter concerning the length of the answer and its nature. The questions of the interviews focused on getting information about the kind of programme employed to teach English at the university concerning the fourth year and the types of questions or exercises involved in examinations.

Thus, the teachers were questioned about the content or titles of the lectures of each module they taught. They were asked about the types of exercises used to consolidate the lectures or the projects students had to present. They were also required to provide the kinds of questions they used in examinations and the way they corrected the students' answers.

The teachers were asked about their method of assessing students' dissertations as well as the difficulties they encountered when evaluating them. They were requested

to state their opinion about their learners' ability to write memoirs and the types of errors produced in their dissertations.

3.6.2- The Questionnaires

A questionnaire is a set of written questions designed by the researcher to be answered by the sample population. It is used to gather two types of data: facts and opinions. Facts refer to those items about which the respondents have knowledge. Opinions are expressions of attitudes or preferences. When the questionnaire asks the respondents about facts, it is called a factual questionnaire. However, it is referred to as an inventory when the informants are required to express attitudes, opinions and preferences (Thomas, 2003).

Most of the time, data about facts and opinions are gathered within the same questionnaire. *"Questionnaires can be designed and used to collect vast quantities of data from a variety of respondents... They can be easily and quickly analysed once completed"* (Wilkinson and Birmingham, 2003: 8). In this way, they help to get a wide range of information.

Questionnaires may include three types of questions. The first type is the close-ended format which requires responses with 'Yes' or 'No', sometimes followed by an additional answer concerning the reason for choosing one of the two options. The second type is the open-ended format which involves questions requiring long answers. The third type of questions is the multiple-choice format in which responses are provided to each question and the informants are requested either to choose only one answer or select several alternatives.

The questionnaires were used to gather data about the students and teachers' opinions concerning examinations especially the ones based on holistic rating. Two types of questionnaires were designed. One was administered to the students and the other was set for the teachers.

3.6.2.1- The Students' Questionnaire

There was an attempt to administer the questionnaire to all the fourth year students. However, not all of them responded mainly because of the factors of attrition represented by the absence of some learners and students' failure to return the questionnaire. Therefore, the accessible population included seventy-two (72) students.

The questionnaire meant to be answered by the fourth year students aimed at collecting information about the learners' way of learning English and coping with examinations. This questionnaire (see the appendix) included nine questions. Four of them were open-ended because they required detailed answers. Three questions were close-ended as they needed to be answered either in a positive or a negative form. Two multiple-choice questions were used to limit the students' answers and make them opt for one alternative.

The first question, which was open-ended, asked the students about the difficulties they encountered when learning English. The second one required them to list the modules they found easy to learn. The third one requested them to list the modules they considered as difficult or requiring much effort. The fourth one, being close-ended, questioned the students about whether they searched for more information in addition to the lectures provided by the teachers or not, urging those who answered by 'yes' to state how they searched for more information and requesting those who responded by 'no' to explain why they did not look for other sources of information and relied only on the lectures. The fifth one, which was also close-ended, asked them if they were influenced by the mother tongue when learning English requesting those who answered in the affirmative to say how their native language interfered in their way of learning English.

The first five questions of the questionnaire had the purpose of getting an idea about the different opinions and views of the students concerning the learning process while the remaining questions aimed at investigating about how the students perceived and handled examinations. Thus, the sixth question which was a multiple choice question, urged the students to state their feeling when required to sit for examinations. Three alternatives were provided making them say whether they felt anxious, afraid or at ease. The seventh one questioned them about the difficulties they faced when answering the questions of the examinations. Being open-ended, it did not restrict the various kinds of responses that could be suggested by the informants.

Question eight was close-ended. It asked the students to state if they liked examinations involving essay writing or not, requesting them to justify their answers by giving the reason why they liked or did not like essay examinations. The last one, being a multiple choice question, inquired about the method they used when answering the questions of the examinations that required from them to write essays. Three options were provided asking them to state whether they preferred to employ their own style, to retrieve the lecture or to combine their style with the one used in the lessons.

The students' questionnaire attempted to collect data about the attitudes and opinions of the fourth year students of English as far as their learning of this language and their way of answering essay examinations were concerned. It also aimed at eliciting the kinds of difficulties they encountered both when learning and when being assessed. The questions involved in this questionnaire were of three types. Open-ended questions were used when it was necessary to obtain the students' views without interfering or anticipating their answers. Close–ended ones were opted for when the question implied the existence of only two possibilities that obliged the informants to answer either in a positive or a negative manner. Multiple-choice questions were also employed in situations where it was compulsory to avoid confusion and limit the informants' answers to those provided as alternatives.

In addition to this questionnaire, another one was designed for the teachers whose opinions are not less important than those of the students.

3.6.2.2- The Teachers' Questionnaire

The questionnaire administered to the teachers tried to provide data concerning their way of preparing and correcting examinations. It was composed of ten questions. Four of them were open-ended leading the teachers to give their answers according to their own views in a free manner. Two close-ended questions were utilized making the informants opt for one possibility among two and supply a short explanation. Four multiple choice questions were used so as to limit the teachers' responses to the provided alternatives from which they had to tick an appropriate answer that corresponded to the case they were encountered with.

The first question asked the teachers about the modules they taught. The second one inquired about the difficulties they encountered when teaching English. The third was a multiple choice question that asked them about the kinds of questions they used in examinations. Three options were presented urging the teachers to make it clear if they employed questions requiring short answers, questions requiring an essay or a combination of questions involving short answers and others requiring a paragraph. It is necessary to mention that this question was opted for in order to find out the frequency or recurrence of using the different types of questions and know the prevailing one. Thus, it did not request the teachers to state the type of examination questions utilized for each specific module as this matter has been raised in the questions of the interviews.

The fifth one, being open-ended, questioned them about the difficulties they found when correcting examinations. The sixth one was a multiple-choice question displaying three alternatives asking the teachers whether when correcting examination papers, they were more interested in form, content or both (form and

content). The seventh question was about the number of scoring points they assigned to form and content. The following one was close-ended requesting the teachers to state if they found it easy to prepare the questions of the examinations and asking them to provide the reason when answering in the affirmative or the negative. The ninth one was an open-ended question inquiring about the kind of mistakes made by most of the students when answering the examination questions. The last one was a multiple choice question that urged the teachers to give their opinion about their students' level by ticking one of the given choices stating if this level was weak, average or good.

This questionnaire aimed at gathering data about how teachers dealt with examinations including the kind of exercises they used and their way of correcting and judging students' answers.

3.6.3- The Tests

Tests refer to a set of mental tasks that are performed by a group of people. They aim at measuring the skills, aptitudes and behaviours.

Tests can be a valuable source of getting information about the learners, the teaching process and the teaching materials. Bachman (1990) suggests the possibility to use tests as tools for research into the nature of human abilities. This means that the researcher may utilize tests not only as a form of measurement of the learners' performances for the sake of making decisions about their success or failure but also as a kind of analysis of the teaching and testing situation.

According to Scott and Morrison, "... *Tests allow researchers to compare the results from their own tests given to a small sample of respondents with the results from a national population*" (2006: 245). They can be useful instruments in research about educational evaluation as they can provide various types of information about the learners' abilities, the testing situation and the factors that may influence or hinder the process of assessment.

They are usually employed as research tools in experiments and surveys. Sometimes, they are utilized in case studies. Tests may be used to furnish qualitative or quantitative data (Thomas, 2003).

In this research work, tests were employed with a group of students who were admitted to the fourth year. They represented a form of research instruments to gather data about how one teacher or different teachers assessed the students' writing performance. They aimed at collecting information about scorer's reliability and making a comparison between the scores awarded by each teacher to the same copy as well as the degree of achievement of the group of students to whom the tests were administered.

A test was administered to thirty (30) students during the period of time of one hour and a half. These learners were required to write an essay in which they discussed how colonization caused the suffering of the colonized people. The purpose of this test was to investigate about intrarater reliability so that to get informed about how the scores given by the same teacher could differ if the scoring procedure was repeated several times. Therefore, the test was corrected three times by only one teacher. A considerable time interval took place between each correction and the other.

Another test was administered to a group of fifteen students who had to write an essay for a period of one hour and a half. They were given three questions and were asked to choose one. The result was that each question was chosen by five students. The use of multiple questions aimed at studying the hidden variable which was the topic effect in order to get informed about the impact of a given essay question on the learner's performance. Three topics were designed. One dealt with civilization. The second concerned literature while the last one focused on linguistics. The first topic required the learners to discuss a statement implying that colonization had serious effects on the colonized countries. The second one involved a discussion of a quotation claiming that the novel was history as it might have

happened. The third question requested the students to discuss a quotation stating that language is a means of communicating ideas and feelings by the use of conventionalized signs.

The learners were given the possibility to state which type of topic they will deal with. Five students chose to answer the question related to civilization. Five students opted for the one concerning literature while the remaining five students selected the question of linguistics. Therefore, three subgroups of students were involved in this test.

The first subgroup was called group A. It included those students answering the question of civilization. The second one was related to the students who chose the topic of literature and was named group B. The third subgroup was referred to as group C. It involved those students responding to the question of linguistics. After the test has been administered, the answers of the students of the three subgroups were photocopied in order to get several duplications of each copy. Then, the papers of each subgroup were given to three teachers to correct them. Thus, the answer of each student from the group was corrected three times by three different teachers. It should be mentioned that the raters who scored the papers of group A were called teacher A1, teacher A2 and teacher A3. The assessors involved in the assessment of the prompts of group B were named teacher B1, teacher B2 and teacher B3. The scorers concerned by the correction of the essays of group C were referred to as teacher C1, teacher C2 and teacher C3.

The sample involved in each subgroup was very small. This was due to two main reasons. First, the correction of a few papers was not time consuming. This avoided creating a burden for the teachers who participated in this task. Second, the main concern was the degree of agreement among independent scorers rather than the students' achievement.

The purpose of such a procedure was to gather information about interrater reliability, so that to know to what extent the grades awarded to each paper would be compatible and also to make a comparison between the differences in the assessors' way of rating the students' essays.

In fact, when requested to correct those papers, there was no restriction imposed on the teachers concerning whether to give more importance to form rather than content or vice versa but they were told to assess the students' answers holistically using their own impressions and points of view. They had, thus, to make a hint at the mistakes whenever possible and give a mark to each paper.

In addition to the tests, document analysis was also employed as a research instrument.

3.6.4- Document Analysis

Documentary research relies on the use of data drawn from the educational materials that are available to the researcher. Although it was a neglected genre, it became to be considered as an essential tool of investigation in the 1980's (Scott and Morrison, 2006). The use of documents in educational research aims at providing a larger data set which assists the researcher in the process of analysis and gives a wider context for information gathering.

Documents can take different forms including diaries, records, biographies, timetables, minutes of meeting, samples of students' work, emails, statistics, archives, artefacts, conversations, newspaper articles, books and public records. They may be written like notes or visual such as photographs. They can be classified into several continua. Thus, they can be anonymous or authored, published or unpublished and professional or lay. They can relate to public (official) or private (personal) domains. They may contain facts or beliefs (Cohen et al., 2007)

The quality of the evidence provided from documents must be based on four criteria which are: authenticity, credibility, representativeness and meaning. Authenticity relates to the genuine and unquestionable origin of the evidence in the sense that the researcher has to ascertain that the document remained untouched, unedited and stored directly after being produced. Credibility refers to accuracy and sincerity which means that the information provided are free from bias and distortion via the use of believable documents. Representativeness implies typicality by revealing if the obtained evidence is typical of its kind and if it is not, the researcher has to investigate about the extent to which its untypicality is known. The criterion of meaning is linked to the degree to which the researcher can make sense of the actual data. These four criteria constitute the features of the validity and reliability of using documents. Nevertheless, the employment of documents in educational research relies mainly on the criteria of representativeness and meaning (Scott and Morrison, 2006).

Documentary analysis has several advantages. First, it helps the researcher to reach inaccessible persons or subjects. It displays little or no reactivity from the part of the writer of the document which leads the investigated phenomenon to be more visible. Large samples can be studied through documents which constitute part of the everyday life of the studied population (Cohen et al., 2007).

In order to be a useful tool, the document has to be studied by employing certain techniques. In fact, the main strategy used by the researcher to analyse documents is the hermeneutic circle which means that the researcher examines the document according to his own frame of reference and attempts later on to conceive the framework of the writer of this document in order to reach a general understanding of the provided information (Scott and Morrison, 2006). This idea is linked to the subjectivity of the reader of the document.

The analysis of documents is described by many researchers as a hermeneutic activity that may have various facets. It can be double hermeneutic in the sense that the researcher interprets a pre-interpreted world which means that he makes

interpretations of documents that have already been ascribed meaning by its writers. This analysis may be triple hermeneutic when the researcher transforms a recorded live event into written data. Thus, the document becomes second hand and is interpreted via another medium embodied in writing which replaces action. In this case, the researcher makes inferences of the primary instrument through the second hand document. The fourth aspect of analysis is quadruple hermeneutic when the document relates to an action that is transformed into a written comment which is analysed by the researcher whose interpretation is examined by a reader (Cohen et al., 2007).

Document analysis is a difficult task mainly because of the issues of inference and bias which often characterize the researcher's interpretations. However, the utilization of documents in educational research is very essential as it enhances the scope of investigation and provides more information that can reflect the practical aspect of the investigated topic. The researcher can make documentary analysis a successful process by eliciting the type of information needed to be extracted from the document and the techniques that have to be employed to achieve this task. This means that a comprehensive analysis of the document should attempt to disentangle facts from the researcher's interpretations.

In educational research, document analysis is a research instrument used for gathering data about learners' performance or program evaluation. It involves the examination of instructional documents which are any written or recorded material such as examination results, syllabi, students' graded papers and lecture notes. This procedure is frequently employed in educational evaluation. *"Document analysis is most often used to enhance and enrich research utilizing other qualitative methods"*(Love, 2003:84).

The analysis of instructional documents helps the researcher to get more information about the field of investigation by directly examining the syllabus, the course content or students' assignment.

> _Document Analysis may be superior to interviewing for collecting some kind of retrospective data... Documents appear more credible than other forms of data, as there is less evaluator bias in the production of such data... Documents are nonreactive._   (Mathison, 2004: 119)

Therefore, this technique has various advantages as it can help to save time and increase the degree of reliability of the process of data collection. Documents are not prepared specifically in response to a request from an inquirer or for some official accounting (Love, 2003). This is why they are an important source of information that enables the researcher to understand the investigated issue and give more details about the real state of the art. Love (2003) identifies various types of documents related to higher education. They include:

a) General documents such as e-mails, websites, flyers and correspondence.

b) Students documents like students' handbooks, class papers, journals and portfolios.

c) Academic documents such as the course schedule and course syllabi.

d) Administration documents like annual reports evaluation, memos and progress reports.

Document analysis may be used to get information about various instructional activities including course practices and students' performance assessment.

In addition to the instruments used for data collection in this research work, document analysis was utilized to provide more information about the learners' writing performance assessment. The documents employed involved students' examination papers and students' marks.

3.6.4.1- The students' Examination Papers

The students' graded papers represented valuable instructional documents that provided more information about the learners' way of handling essays.

The analysis of the students' examination papers was very useful in providing data concerning the types of mistakes produced by the fourth year students. A sort of brief summary of the most common and frequent errors repeated in the examinations of all the modules was supplied. Then, a few samples of students' answers in certain modules were studied to give an idea to the reader about the degree of the learners' ability to write essays taking into consideration both the content and the form of the produced writings. The study of such papers also made it possible to spot certain mistakes that may be seen as unacceptable errors that had rather not to be made by fourth year students.

The aim from studying the students' examination papers was to draw a link between the theoretical assumption concerning the utilization of holistic rating for testing writing and the real state of such an assessment by giving a picture of the kind of production achieved by the testees and the way the teachers coped with such a situation.

Samples of students' answers in the examinations of the modules of African civilization, literature, psychopedagogy and TEFL were used in order to check if the students performed differently when writing essays in these different modules taking into account the fact that the module of literature was introduced in the second year; that of psychopedagogy was dealt with in the third year while the module of African civilization and TEFL have been introduced only in the fourth year. This sort of study enabled to discuss whether there was a problem of content rather than form or vice versa.

The sample of the examination papers included forty (40) copies in the module of civilization, ten (10) copies in the module of psychology, ten (10) copies in the module of literature and ten papers dealing with the module of TEFL. The analysis of these papers focused on the teachers' way of spotting the learners'

mistakes, the students' methods of responding and the types of mistakes they produced in their essays.

3.6.4.2- The Students' Marks

The study of the students' grades obtained during the year and the kind of individual differences they displayed was necessary to provide information about the extent to which these learners were able to perform well in essay examinations. It also made it possible to draw certain conclusions about the testing situation by examining the degree of differences existing between the students' marks.

A sample consisting of the marks of thirty eight (38) students in the module of African civilization was analysed to elicit the differences in the learners' writing performance in the various examinations administered during the year.
Another sample of the marks of a very small group composed of five students has been opted for to try to make a comparison between these students' achievement in the first and the second examination so that to conclude if they have improved or not. Finally, the general results of the examinations of all the modules were discussed to infer the extent of similarity or difference in the learners' writing performance in the different modules.

The study of the learners' marks was meant to provide useful information about the degree of the learners' performance in examinations that involved essay questions as well as the kind of decrease or increase of their level from a semester to another.

3.7- Conclusion

This chapter has been devoted to the description of the main instruments utilized to undergo this research work which focused on employing interviews, questionnaires, tests and document analysis as means of gathering information.

It was necessary to present and describe the research tools so that to give an idea about the way of collecting information to achieve the aims and objectives of the investigation taking place. Thus, some instruments have been chosen mainly because the subject dealt with holistic scoring. For instance, certain materials such as tests and document analysis were more helpful and practical than observing the learners.

As the case study was related to assessment, it was essential to examine the marks of the students and the essays they produced in the examinations so that to be able to decide about the kind of difficulties faced by both the learners and the teachers when dealing with such a type of assessment.

The research tools described in this chapter made it possible to collect information about the case of the assessment based on holistic rating which is confined to essay examinations. The gathered data will be analysed and interpreted in the next chapter.

CHAPTER FOUR:

DATA ANALYSIS
AND INTERPRETATION

CHAPTER FOUR:

DATA ANALYSIS AND INTERPRETATION

4.1- Introduction

This chapter provides an analysis of the information gathered from the interviews, the questionnaires, the tests, the examinations' papers and the students' marks. It also presents an interpretation of the results obtained using as a basis the statistics available to supply the arguments with tables and figures to make the explanations clearer.

Finally, the main results and conclusions drawn from this study are summarized in this chapter to state to what extent the questions raised by this research work have been answered and make a correlation between the presented hypotheses and the real state of the art that is pictured in the obtained results.

4.2- The Analysis of the Interviews

The interviews revealed that the majority of the teachers taught several modules related to more than one discipline. Concerning the programme they employed, all of them agreed that, in general, they followed the broad guidelines put forward by the Ministry of Higher Education but in reality the lectures they prepared were basically produced according to their own research and endeavour to meet the learners' needs in each level.

Those teachers who teach grammar and phonetics focus on dividing the lesson into theoretical parts and practical ones involving exercises. The ones who are in charge of the module of linguistics for the first and second year try to backup the lessons with a few exercises whenever it is possible to do so. The teachers in charge of oral expression tend to urge the learners to make dialogues, discussions and plays of their own production. Those who teach civilization, literature, psychopedagogy, linguistics and TEFL give lectures that supply the students with general information about the taught subject urging them to consolidate the given data with further reading and requesting them to present projects either individually or in small groups specially in the modules of civilization and literature.

As far as the testing process was concerned, the teachers affirmed that the types of questions used depended on the nature of the taught module. The following table provides an idea about the various kinds of questions used in the module of phonetics as well as the modules related to language skills including grammar, written expression and reading.

Modules	*Types of examination questions*
Grammar	Exercises requiring the students : - to write verbs correctly - To form nouns, verbs or adjectives - To supply articles or propositions - To transform from passive to active or vice versa - To report speech
Written Expression	Exercises urging the testees: - To supply punctuation and capitalization - To find the type of sentences - To write either simple, compound or complex sentences - To find the type and function of the subordinate clauses - To write a short paragraph about a certain topic. - To write a letter
Reading	- A set of comprehension questions - An exercise requiring the learners to write a short paragraph
Phonetics	- Four or five direct questions about the speech organs, the production of sounds, consonants and vowels - An exercise or two dealing with transcriptions

Table 4.1- Types of questions used in the examinations of the modules related to language skills

From the above table, it is clear that the teachers use a variety of exercises in the examinations of grammar, written expression, reading and phonetics so that to assess the learners' mastery of the different elements studied taking into account both the theoretical and practical side. In the examination of oral expression, the teachers employ questions dealing with different topics that should be discussed orally by the learners.

Concerning the examination of linguistics, the questions used in the first and second year involve a statement to comment on in addition to an exercise requiring definitions of studied concepts. In the third and fourth year, the examinations require the learners to comment on a direct question or to discuss a quotation in an essay of a considerable length.

In the examination of the module of British civilization, a question involving a statement or a quotation to comment on is employed. For the module of American civilization, one teacher requires the students to comment on a given statement while the other gives them a direct question. In the examination of American literature, the students have to write an essay on one of two given topics. The first topic requires them to comment on a quotation speaking about one of the lectures whereas the second one involves analysing a short passage taken from the studied novels or short stories by commenting on the style, the setting, the plot, the characters and the literary devices used.

As far as the module of British literature is concerned, the students of the second year are assessed in the first semester by being asked to study a given poem at the level of main ideas, literary genre, assonance, consonance, alliteration and verse. In the second semester, they are requested to choose one of two topics which are in the form of a direct question or a statement to comment on in an essay. For the third year, the questions involve a commentary on a statement or a direct question.

The examination of psychopedagogy includes either four or two exercises. One of them embodies a direct question to be answered in 10 or 12 lines. The others involve statements to be explained by examples. Third world literature is assessed through the discussion of a statement, the explanation of a poem or the analysis of a

passage from a short story. For the module of African civilization and TEFL, the learners are asked to comment on a direct question or a quotation.

Generally speaking, the majority of the modules of the first year involve an assessment based on analytic rating whereas most of the second year modules and all those of the third and fourth year are assessed using holistic rating.

Concerning their method of assessing students' dissertations, the teachers affirmed that they relied on their overall impression and general opinion about the written work. Some of them stated that they gave more importance to the style rather than the content while others focused on examining the methodology employed, language mastery and the content of the memoir. One teacher added that oral presentation was also taken into consideration.

Regarding the difficulties they encountered when evaluating dissertations, the teachers said that they found problems in assessing memoirs because of their subjective judgement and their sympathy towards the learners they supervised.

As far as the students' ability to write projects was concerned, the teachers affirmed that the majority of the students were unable to produce excellent dissertations as they had difficulties in mastering the target language. Among the errors made by the students in their projects, the teachers listed the lack of proofreading and the need of methodology. They added that the learners produced certain ambiguous sentences and grammatical mistakes. Also, some teachers claimed that the learners tended to focus on the literature review more than the analysis of the problematics which led their projects to appear more theoretical than practical.

4.3- The Analysis of the Students' Questionnaire

The questionnaire designed to the fourth year learners was answered by 72 students only. When questioned about the problems they faced in learning English, the students provided various responses. The majority of them said that they found difficulties in pronunciation and understanding the teachers. Most of them affirmed that there was a lack of books. Some of them stated that they did not have sufficient

knowledge of English grammar and vocabulary. Others could not write well using good English and found the lectures difficult. Some of the students affirmed that they lacked methodology in writing essays. A few of them believed that they had problems in learning English because of the lack of practice as this language was not spoken outside the classroom.

The following bar graph shows a classification of the students' learning problems according to the number of times they were listed by the informants.
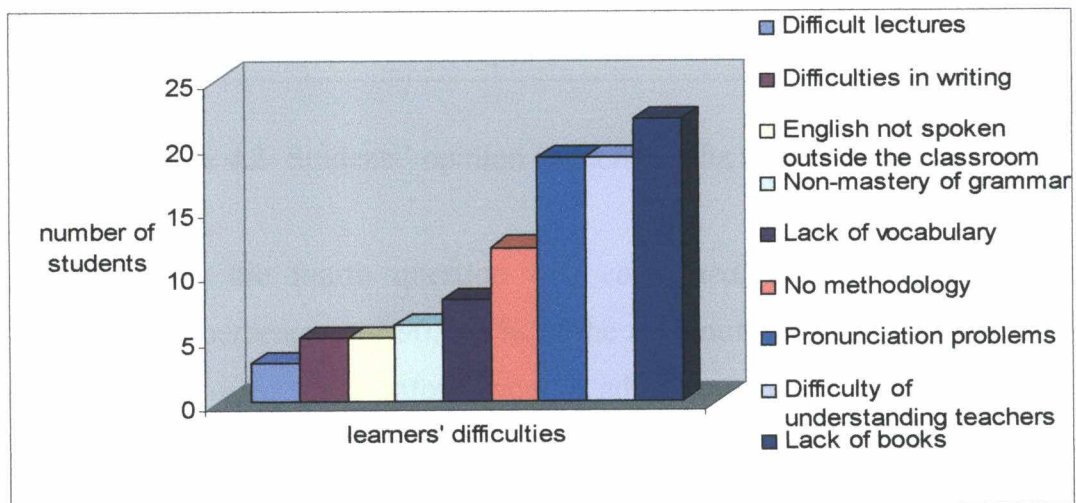
Bar graph 4.1- Students' difficulties in learning English

From the above bar graph, it can be understood that the most important difficulties are the lack of books, problems of pronunciation and the difficulty of understanding the teachers.

Concerning the modules they found easy to learn, the students gave different answers. Most of them considered literature and psychopedagogy as being easy. When answering the third question, the majority of the learners agreed on the point that the modules of linguistics and TEFL were the most difficult to learn. For civilization, it can be stated that some of them found it easy while others thought it hard to study.

The following table summarizes the responses of the second and third question.

Regarding the sixth question, the majority of the students said that they felt anxious when required to sit for an examination. The following table gives an idea about the different responses of the informants.

Feeling	Number of students	percentage
Anxious	45	62,50 %
Afraid	17	23,61 %
At ease	10	13,88 %

Table 4.3- Students' feeling when sitting for examinations

Therefore, most of the respondents have a feeling of anxiety when sitting for examinations. Some of them feel afraid while just a few of them find themselves at ease.

When asked to list the difficulties they faced when answering the questions of the examinations, three (03) students, representing a percentage of 04,16 %, did not answer, while the others gave different responses which are summarized in the following table :

Students' problems	Number of students	percentage
- The difficulty of writing essays	16	22,22 %
- Inability to manage time	12	16,66 %
- The problem of methodology	08	11,11 %
- Incapacity to understand the question of the examination	08	11,11 %
- Inability to organize ideas	07	09,72 %
- Grammatical and spelling mistakes	07	09,72 %
- Lack of concentration	06	08,33 %
- Lack of vocabulary	05	06,94 %

Table 4.4- Students' difficulties in examinations

From the above table, it can be inferred that most of the students face difficulties in writing essays. Some of them are unable to manage time as they can not finish their essay and fully provide an answer within the period of time designed for the examination. Other students have a problem of methodology concerning the way to answer questions or comment on quotations. Another group of learners can not understand the examination questions and do not totally grasp what is required from them. Some students are unable to organize their ideas in a logical way while others make grammatical and spelling mistakes. A few of them have a problem of concentration as they can not think only about the answer and worry about the mark which leads them to forget a great amount of the information they learnt. Other students have a lack of vocabulary which makes them unable to express themselves freely.

Concerning question eight, the majority of the informants answered by 'yes'. Thus, forty-five (45) students, representing the percentage of 62,50 % out of the total number of respondents, said that they liked having examinations that required from them to write an essay. On the other hand, twenty-seven (27) students, representing 37,50 %, answered by 'no' asserting that they did not like essay examinations. The following pie chart describes the students' responses.
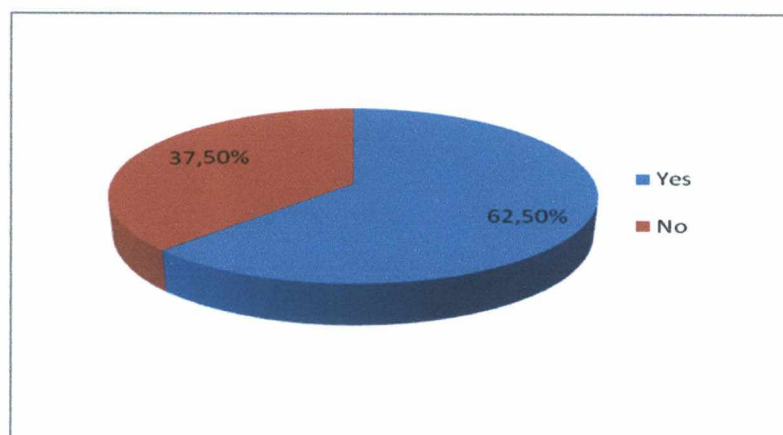
Pie chart 4.1 – Students' attitudes towards essay examinations

When required to justify their answers, both categories of students gave different arguments. The first group affirmed that essay examinations provided the learners with an opportunity to use their own style, to improve their writing skill, and to display their knowledge of the target language.

The second category stated that they disliked such examinations because of the difficulty of essay writing, the non mastery of the writing skill and their lack of vocabulary.

The table below gives an idea about the various arguments given by the students.

Liking essay examination			Disliking essay examinations		
Arguments	Number of students	Percentage	Arguments	Number of students	Percentage
- Using one's own style	32	44,44 %	- Writing an essay is a difficult task	17	23,61 %
- Improving one's writing abilities	04	05,55 %	- Non-mastery of the writing skill	04	05,55%
- Acquiring a method of analyzing and discussing	04	05,55 %	- No methodology of writing essays	03	04,16 %
- Showing one's knowledge of the English language	05	06,94 %	- Lack of vocabulary	03	04,16 %

Table 4.5- Students' arguments concerning their like or dislike of essay examinations

Therefore, most of the students like essay examinations because they enable them to express themselves freely.

As far as the last question of the questionnaire was concerned, the majority of the students said that when answering examination questions, they used their own

style with the one employed in the lesson. Their responses are illustrated in the following pie chart.
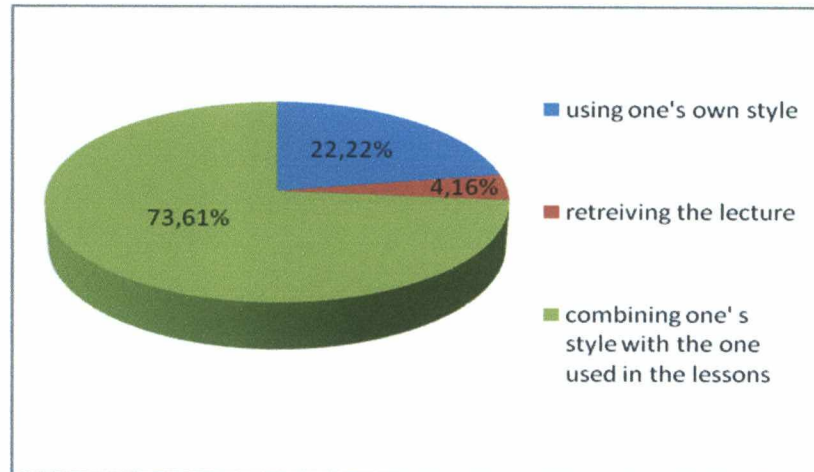
Pie chart 4.2 – The students' way of answering essay examinations

Thus, three (03) students, representing 04,16 % of the total number of informants stated that they retrieved the lecture while sixteen (16) students, representing 22,22 % of the total number of respondents, said that they utilized their own style. The remaining fifty-three (53) students, representing the percentage of 73,61 %, claimed that they preferred to combine their style with the one used in the lessons.

4.4. The Analysis of the Teachers' Questionnaire

The questionnaire designed to the teachers was answered by twelve (12) teachers. Concerning the answer of the first question, it can be stated that the majority of the teachers are in charge of two or three modules involving either a language skill or a content area.

The following table summarizes the responses of these teachers.

Modules \ Teachers	A	B	C	D	E	F	G	H	I	J	K	L
Reading						X						X
Grammar		X								X	X	
Written expression	X			X				X		X		
Oral expression						X	X					
Phonetics									X		X	
Linguistics		X				X	X		X			X
American civilization	X											
American literature					X							
British literature				X				X				
psychopedagogy			X									
TEFL						X	X					X

Table 4.6- The modules taught by each teacher

From the above table, it can be noticed that only one teacher is in charge of modules related just to language skill. Each one of the remaining teachers is in charge of , at least, one module related to language skills and one or two modules linked to content knowledge.

Regarding the difficulties encountered when teaching English, eight (08) teachers, representing 66,66 % of the total number of the respondents, agreed on the point that they had problems in explaining the lectures due to the students' weaknesses which were related to their language proficiency level. Four teachers, representing 33,33 %, stated that the learners were not motivated and did not participate in the classroom which led the teachers to feel a lack of feedback and collaboration. Two of them added that they had difficulties in teaching because of the over crowded groups as well as the lack of laboratories which caused a shortfall of practical sessions.

Concerning the main purpose of teaching any module, the teachers provided different answers involving a specific percentage given to each purpose on the basis of its priority. Their responses can be classified into five groups. Each group represents an answer of a number of teachers who agreed on the same percentage referring to the degree of priority of the mentioned teaching objective. Their responses are shown in the table below.

Purpose of teaching Groups	First group	Second group	Third group	Fourth group	Fifth group
Students' mastery of the English language	70 %	50 %	40 %	30 %	10 %
Students' mastery of a specific knowledge related to the taught module	30 %	50%	60 %	70 %	90 %

Table 4.7- The percentages provided by the teachers concerning the purposes of teaching

Therefore, one teacher stated that more priority was directed to the mastery of English, whereas four teachers affirmed that they were interested in making the students master both the English language and the knowledge related to the module. On the other hand, two teachers claimed that they gave a little priority to the mastery of the knowledge related to the module while four teachers said that they were more concerned with teaching the content of the module.

Finally, one teacher stated that the students' mastery of the knowledge related to the module was the main purpose while the mastery of English was given little attention in the modules that did not involve grammar and written expression.

The following table provides the number of teachers opting for the different responses explained in the above table:

Groups	First group	Second group	Third group	Fourth group	Fifth group
Number of respondents	01	04	02	04	01
Percentage	08,33 %	33,33 %	16,66 %	33,33 %	08,33 %

Table 4.8 - The teachers' responses to the third question.

Therefore, the teachers' perceptions of the objectives related to the teaching of any given module are not the same. A minority opts for the extremes which means that they either focus on the mastery of the target language or the mastery of the content of the module. On the other hand, most of the teachers give equal importance to the mastery of the language and the content knowledge of the module while certain teachers focus just on the students' mastery of the content knowledge.

When required to mention the kinds of questions they employed in the examinations, the teachers provided different alternatives according to the nature of the module. The following table explains their answers.

Modules / Types of questions	Phonetics	Grammar	Written Expression	The modules related to content areas
Questions requiring short answers	X			
Questions requiring an essay				X
A combination of short answers and a paragraph		X	X	

Table 4.9- Types of questions used in examinations

Thus, the majority of the modules including civilization, literature, psychology, linguistics and TEFL involve examinations requiring an answer in the form of an essay since these modules are related to content areas or disciplines.

Concerning the difficulties they found in correcting the examinations, the teachers stated that they faced problems in assessing the students' papers because of the great number of grammatical and spelling mistakes made by the learners as well as the problem of the coherence and cohesion of the essays.

Also, some teachers noted that the answers were sometimes irrelevant, out of subject or including long unnecessary details. Others said that some students had a bad handwriting which led the teachers to waste more time in trying to understand what was written. In addition to this, the teachers affirmed that they found it difficult to cope with the students' long answers that were full of mistakes as they had to give back the marks on time despite the large number of copies.

When asked whether they were more interested in the form or the content of the essay, eleven (11) teachers representing 91,66 % of the total number of the respondents, said that they gave importance to form and content when they corrected examinations. However, one (01) teacher, representing 08,33 % of the total number, claimed that he was more interested in correcting the content of the examination answer.

The teachers' responses concerning the number of grades assigned to the form and the content of the essay are classified into five groups. Each group represents a specific response. These answers are summarized in the following table.

Grades Form or content	Number of points (out of twenty)				
	First group	Second group	Third group	Fourth group	Fifth group
Form	10	08	12	15	00
Content	10	12	08	05	20

Table 4.10 – The grades assigned to the form and content of essays.

Therefore, a group of teachers awards an equal number of grades to both the form and the content of the essay. The second group gives more importance to content while the third group focuses on correcting the form. The fourth group gives little importance to the content while the fifth one totally ignores the form and relies on correcting the content of the learners' answers.

The table below provides a classification of the number of teachers corresponding to each group of responses.

Groups Teachers	First group	Second group	Third group	Fourth group	Fifth group
Number of respondents	05	04	01	01	01
Percentage	41,66 %	33,33 %	08,33 %	08,33 %	08,33 %

Table 4.11- Teachers' responses concerning essay scoring.

Therefore, a considerable number of the questioned teachers assign the grade of ten out of twenty (10/20) to the form of the essay and the same grade is designed to the content. However, another group which represents one third of the respondents awards the grade of eight out of twenty (08/20) to the form. Two teachers focus on the form while one teacher gives all the priority to the content.

Regarding the answer of the question requiring to state if they found it easy to prepare the questions of the examinations, nine (09) teachers representing 75 % of the total number of the respondents, replied by 'yes' arguing that they found it easy to prepare the examination tasks because the questions were related to the content of the courses. On the other hand, three (03) teachers representing 25 % of the respondents affirmed that it was difficult to prepare the questions of the examinations because they had to think about topics that required careful reflection rather than automatic answers. Also, they had to take into account the level of the students and employ questions that aimed at enhancing process rather than product writing. The teachers' answers are summarized in the following pie chart.
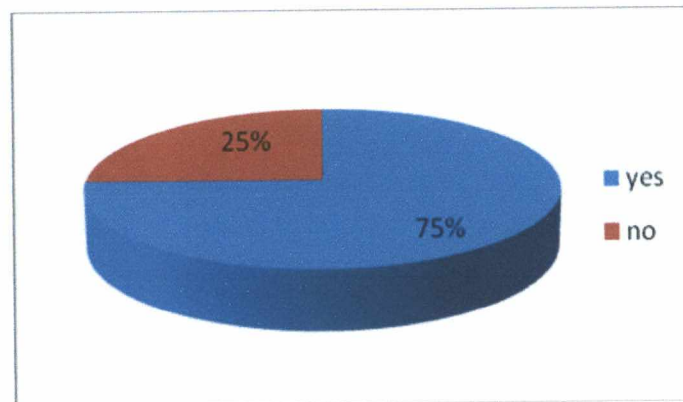
Pie chart 4.3- Teachers' opinions about tests' construction

From the pie chart, one can infer that the majority of the teachers think that it is easy to prepare the questions of examinations since they turn around the content of the lectures.

When questioned about the types of mistakes made by the students when answering the examination questions, all the teachers said that the students produced grammatical and spelling mistakes. They also mentioned the lack of cohesion within the learners' essays. They added that some students wrote too lengthy essays that exceeded the normal length.

Concerning their opinion about the students' level, five (05) teachers, representing 41,66 % of the total number of the respondents, affirmed that the students had an average level. However, seven (07) teachers, representing 58,33 % of the respondents, stated that the students' level was weak.

4.5- The Analysis of the Tests' Results

The analysis of the marks assigned by the same teacher to the sample comprising thirty (30) students revealed a degree of inconsistency in the scores awarded in different occasions. The following table provides an illustration of the marks awarded to the same papers by one teacher.

Students	First mark	Second mark	Third mark	Range
S1	9	9	8	1
S2	10	9	9	1
S3	9	8	9	1
S4	9	8	9	1
S5	12	13	11	2
S6	11	12	12	1
S7	8	9	10	2
S8	7	9	9	2
S9	10	12	10	2
S10	7	8	8	1
S11	9	10	10	1
S12	8	6	8	2
S13	10	10	9	1
S14	7	8	8	1
S15	8	5	6	3
S16	9	10	9	1
S17	13	12	10	3
S18	6	8	8	2
S19	9	10	9	1
S20	10	11	10	1
S21	10	10	10	0
S22	10	11	10	1
S23	10	9	8	2
S24	8	8	7	1
S25	12	13	12	1
S26	12	11	10	2
S27	7	8	8	1
S28	9	8	9	1
S29	9	7	9	2
S30	10	11	9	2

	First mark	Second mark	Third mark
The Mean	09,27	09,43	09,13
Standard Deviation	1,65	1,93	1,28

Table 4.12- Students' scores awarded by the same teacher

The examination of the papers that reflected inconsistent scores revealed that this inconsistency was due to content factors which referred to the cases when the essays involved ambiguous ideas or language factors which means that the compositions included a variety of grammatical mistakes and irrelevant use of vocabulary. At certain times, both factors were apparent. This means that such factors often distort the assessor's judgement and lead the scores to fluctuate.

Concerning the second test, it can be stated that there were certain distinctions in the scoring of the students' essays within the same group or from a group to another. The first group of teachers who corrected the essays of group A employed different ways of spotting learners' mistakes. Two teachers just underlined the mistakes found in the essays while the third one underlined the mistakes and corrected them. The marks awarded to the various prompts are elicited in the following table.

Teachers / Students' papers	Teacher A1	Teacher A2	Teacher A3	Range
P1	12	11	10	2
P2	12	10	08	4
P3	13	11	11	2
P4	11	09	10	2
P5	10	07	07	3

The Mean	11,6	9,6	9,2
Standard Deviation	1,02	1,5	1,47

Table 4.13- The marks of the students of the first group

From the above table, it can be stated that the papers of group A were given approximate grades by teacher A2 and teacher A3. There is a slight difference between the marks awarded by teacher A1 and those of teacher A2 and A3 with the

exception of the papers (P2) and (P5) which display a large difference. The range of the scores of three papers equals two (2) points. The range of the scores of one paper is three points and that of another is four points.

The comparison of the scores awarded by teacher A1 and teacher A2 leads to the conclusion that there is a high degree of inconsistency between the scores. The same remark can be stated concerning the scores assigned by teacher A1 and A3. In both cases, just one paper out of five obtained consistent scores. When comparing the ratings of teacher A2 and A3, one can confirm that the scores of four papers out of five are quite reliable. The mean of the marks given by teacher A1 is 11,6 while the means of the scores assigned by teacher A2 and A3 respectively equal 9,60 and 9,20. The difference between the means of teacher A1 and teacher A2 is two (2) points. The difference between the means of teacher A1 and teacher A3 equals 2,40 which is about two points and a half. This means that there is an extent of variability between the scores assigned by teacher A1 and those scores awarded by the two other teachers. On the other hand, the difference between the means of teacher A2 and teacher A3 is 0.40 which is about half a scoring point. Therefore, there is a high degree of interreliability in the scoring of teacher A2 and teacher A3.

The standard deviations of the scores of teacher A1, teacher A2 and teacher A3 respectively equal (1,02), (1,5) and (1,47). This means that the scores awarded by teacher A1 are slightly closer to the mean than the scores of teacher A2 and teacher A3. The standard deviation of the marks awarded by teacher A1 reveals that there is a certain inconsistency between these scores and those assigned by the other two teachers. However, the standard deviation of the scores of teacher A2 is approximately equal to that of the marks of teacher A3. Thus, the scores awarded by these two teachers are reliable.

Regarding group B, the teachers corrected the mistakes in distinct manners. Teacher B1 circled the spelling and grammar mistakes and corrected the other types like punctuation and capitalization. When some words were missing, she supplied them. Teacher B2 underlined the mistakes and used some symbols like 'G' to refer to grammar, 'rep' to mention repetition and 'X' to state that a word was missing.

Teacher B3 underlined the mistakes and wrote some notes on the margin to refer to the types of mistakes. The following table illustrates the marks awarded by these teachers to the students of this group.

Teachers / Students' papers	Teacher B1	Teacher B2	Teacher B3	Range
P6	8	6	9	3
P 7	7	9	6	3
P 8	6	7	7	1
P 9	8	10	8	2
P 10	11	13	11	2

The Mean	8	9	8,2
Standard Deviation	1,67	2,45	1,72

Table 4.14- The marks of the students of the second group

It can be noticed from the above table that teacher B1 and B3 assigned the same grades to the papers P9 and P10 while they gave approximate marks to P6 and P8. However, teacher B2 diverged from the other teachers mainly in grading the papers P6 and P7. She agreed with the teacher B3 on the same mark concerning the paper P8.

The range of the scores of one paper is one point; two papers have a range of two points; the range of two papers is three points. When comparing the marks awarded by teacher B1 and teacher B2, one can notice that the scores of one paper out of five are fairly consistent. The same remark can be made concerning the scores of teacher B2 and B3. Regarding the scores of teacher B1 and B3, it can be stated that the marks of all the five papers are fairly consistent. This means that there is a degree of interreliability in the scoring of teacher B1and teacher B3.

The means of the scores of teacher B1, B2 and B3 are respectively (8), (9) and (8,2) which means that the difference between the means of teacher B1 and B2 is one point while the difference between the mean of teacher B2 and B3 equals 0,80

which refers to the four fifths ($\frac{4}{5}$) of a scoring point. Therefore, there is an extent of variability between the scores of teacher B2 and those scores of teacher B1 and B3. The difference between the means of teacher B1 and B3 is 0,20 which is the fifth ($\frac{1}{5}$) of a scoring point. Thus, the scores of these two teachers are approximate to each other.

The standard deviation of the scores of the three teachers respectively equal (1,67) ,(2,45) and (1,72). The difference between the standard deviation of teacher B2 and B1 is (0,78); that of teacher B2 and B3 is (0,73). This implies that there is a variation between the scores of teacher B2 and those of teacher B1 and B3. On the other hand, the difference between the standard deviation of teacher B1 and teacher B3 equals (0,05). As a result, there is a degree of consistency in the scoring of teacher B1 and teacher B3.

As far as group C was concerned, two teachers spot the students' mistakes by circling them while the third one underlined them. The table below gives an idea about the marks assigned by these teachers to this group.

Teachers / Students' papers	Teacher C1	Teacher C2	Teacher C3	Range
P11	09	10	08	2
P12	08	09	10	2
P13	08	09	07	2
P14	07	06	09	3
P15	12	10	13	3

The Mean	8,80	8,80	9,40
Standard Deviation	1,72	1,46	2,05

Table 4.15 - The marks of the students of the third group

From the above table, it can be inferred that the range of three papers is two points while two papers have a range of three points. When comparing the scoring of teacher C1 and teacher C2, it can be noticed that the marks of four papers out of five are fairly consistent. Regarding the scores of teacher C1 and teacher C3, one can say that the marks of two papers out of five are consistent. The same remark is made concerning the scores of teacher C2 and teacher C3. The mean of the scores of teacher C1 equals (8,80); the scores of teacher C2 have the same mean. Therefore, there is a degree of consistency in the scoring of these two teachers. The mean of the scores of teacher C3 is (9,40); the difference between the mean of teacher C3 and those of teacher C1 and teacher C2 equals (0,60) which is more than half a point.

The standard deviation of the scores of teacher C1, teacher C2 and teacher C3 are (1,72), (1,46) and (2,05) respectively. The difference between the standard deviation of teacher C3 and teacher C1 is (0,33); that of teacher C3 and teacher C2 is (0,59). Therefore, the scores of teacher C3 vary from those of teacher C1 and teacher C2. However, the difference between the standard deviation of teacher C1 and teacher C2 equals (0,26). This means that there is an extent of interreliability in the scoring of these two teachers.

Generally speaking, in each group, scoring consistency exists between two teachers while the scores of one teacher display a sort of variability. This means that the scores of the two thirds ($\frac{2}{3}$) of the group are reliable. This number represents 66,66% of the total group of teachers. One can conclude that some teachers perceive students' writing in a similar manner. Although there is not an exact scale, there is an extent of reliability in scoring

Concerning the correction of the learners' mistakes by these teachers, one can state that there are certain distinctions in the form of the corrections employed but the purpose remains the same which is to remind and warn the students of the types of mistakes they are making. The following table provides the different ways of correcting such mistakes.

Group of teachers / Method of mistakes' correction	Group A			Group B			Group C		
	A1	A2	A3	B1	B2	B3	C1	C2	C3
Underlining	X	X							
Underlining and correcting			X						
Underlining and using symbols					X				
Underlining and using notes						X			
Circling mistakes							X	X	
Circling mistakes and correcting				X					X

Table 4.16 - The teachers' ways of correcting students' mistakes.

One can infer that three teachers, representing 33,33% of the total number of teachers involved in the test's correction, mention the mistakes and correct them while four of them, representing 44,44%, just spot the mistakes by underlining or circling them. However, two teachers, representing 22,22%, refer to the type of the mistakes by underlining and using symbols or notes.

As far as the kinds of mistakes made by the students are concerned, one can say that the difference in the topics of the essay questions did not affect the learners' way of writing. Most of the students of the three groups had a lack of knowledge concerning the subject and faced difficulties when writing as they made mistakes related to punctuation, capitalization, spelling, and grammar. An illustration of these kinds of mistakes is given in the table below.

Type of mistakes	Examples from the students' essays
Capitalization	* charles dickens – bleak house
Punctuation and capitalization missing	* other colonies might affected a big development; Socially economicly...but After the colonization, it developed ..., it changed...
Incorrect spelling	* Religious – ilegal – bariere – exemple – pauverty – peopel * feilds – interfeare – semmantics – defferent – ourselfes.
Wrong use of tenses	* Can mentioned - he success – everyone know * The novel provide – to wrote – events has
Wrong word derivation	* The novel is a narrator / using the critic * he should success
Words misused	* ...practising in the war / ...which made the economy lazy * they enforced their language as they did in Algeria
Ambiguous sentences	* The novel provides aesthetic pleasure. * Language regards as a good behaviour. * Colonization helped people to be intelligent and jealousy to their country.

Table 4.17 - Examples of the types of mistakes made by the tested students

It can be stated that the majority of this group of students face certain difficulties in writing essays not only at the level of the form but also at the level of the content of their pieces of writing.

Concerning the content of the essay, one can state that there was a distinction in the learners' perception of the topic. As far as group A was concerned, two students lacked knowledge about the topic of colonialism; one student provided few ideas while two learners had a very good knowledge about the given subject. The same remark can be drawn about group B since two students had sufficient knowledge about literature; two students lacked knowledge and one student gave few details but could not give more explanation. In group C, three students lacked

knowledge about the topic whereas two learners had a good knowledge about the meaning of language.

Thus, the learners seem to have approximately the same attitude towards certain modules as there is not a large degree of variability in their performance when dealing with civilization and literature. However, these students display a difficulty in mastering the content of the module of linguistics.

As far as the extent of the mastery of the content knowledge of all the fifteen students is concerned, six students, representing 40%, have a good knowledge of the topic; two students, representing 13.33%, have a limited knowledge of the given subject while seven students, representing 46,66%, do not possess sufficient information about the topic. The learners' degree of mastery of the content knowledge is provided in the following pie chart.
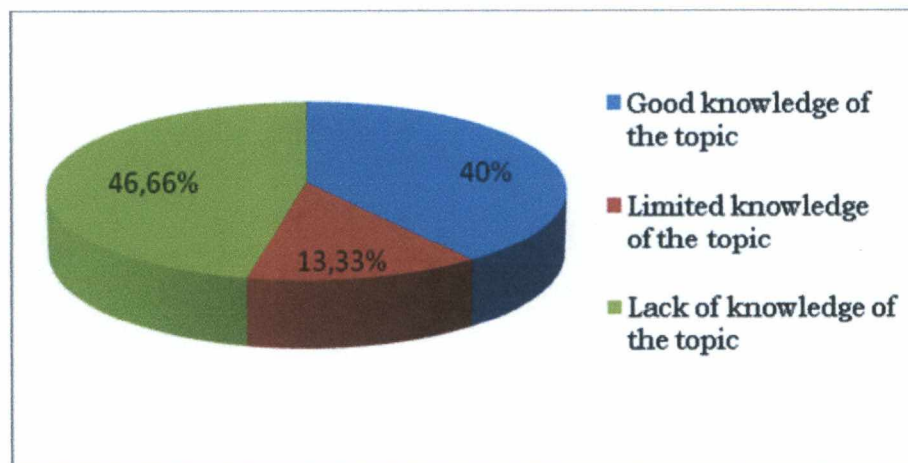
Pie chart 4.4 – Students' content knowledge

Regarding the mastery of language and content knowledge, seven (7) students, representing 46.66% of the total number of this group of learners, had a lack of knowledge of the subject's content and did not master the target language. Three (3) students, representing 20%, possessed good knowledge of the topic but did not master the language. Two (2) students, representing 13.33%, displayed an acceptable level of language mastery but their information about the given topic was restricted to few general details. Three (3) students, representing 20%, had a good command of language and fully managed to discuss the topic. The pie chart below

gives an illustration of the variation in the learners' degree of mastery of language and content knowledge.
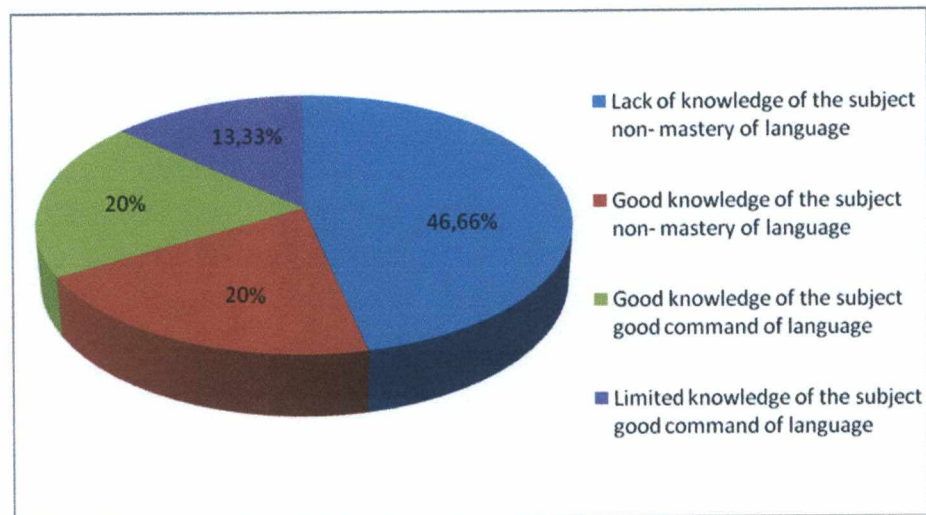
Pie chart 4.5 – The variation in the learners' mastery of language and content knowledge

Therefore, the learners' writing performance may be affected by several factors that are related not only to the students' understanding of the topic and their capacity to convey the required information but also their mastery of the language including both the syntactic and semantic levels.

4.6- The Analysis of the Examination Papers

The study of students' examination papers was necessary to gather information about the types of mistakes made by the learners as well as the teachers' way of correcting these papers.

The way of correcting the examination papers differed from a teacher to another. Two teachers circled the spelling mistakes and provided the mark. One teacher relied on reading the papers and assigning the mark without referring to the mistakes or giving remarks. The other teacher corrected the papers by circling the mistakes, mentioning their type, awarding the grade and writing general remarks about the essay.

From the analysis of all the examination papers, it was noticed that the same type of mistakes was repeated in some of the learners' essays. Thus, the students

mainly made spelling and grammatical mistakes in addition to certain meaning related errors.

The main types of mistakes are summarized in the following table.

Types of mistakes	Erroneous sentences
Capitalization and punctuation missing	* Heart of darkness * he describes…darkness * it is a symbol of darkness * concerning characterization; there are two main characters * Strategies vary, each of us has…. * styles are general characteristics for example you might be more oriented * they were defeated as a result their resistance failed
Spelling mistakes	* Approch * Cognetive * Controle * descrimination * methaphor * Execlussive * beleives * acheivement * availability * diffeculty * Gramatical * devided * existance * pocess * resistence
Grammatical mistakes	* He brought up by his uncle * The story take place * He done bad things * a very important variables * It would be appeared * a specific methods * In the other hand * To answer to this question * There are many method * this factors * The Africans insists * Conrad was successes in writing
Ambiguous sentences	* The text is full of symbolism to make it metaphorically being read and that to understand the meaning is outside. * Concerning style is a term used to rather enduring tendencies * This is a process of educating Christianity * Each of us has a whole host of possible way * Africa was faithfulness * Slavery which defeat African families * Cognitive domain emerge to learning styles * Concerning the learner style that comes naturally to them it might be.

Table 4.18 -The types of mistakes produced by some students.

In fact, the majority of the students approximately made the same kind of mistakes which can be classified into four categories including punctuation, spelling, grammar and meaning-related mistakes. Thus, some students do not perform well in all the modules mainly because of their non-mastery of the language as they roughly produce the same types of mistakes in all the examinations. Concerning the content, they often retrieve the lectures provided by the teachers.

The students' examination papers in the module of African civilization revealed that the majority of learners produced similar mistakes often related to spelling and grammar. The essay question required the students to speak about the different African reactions towards colonialism. The learners' responses varied. The majority of them were able to discuss the topic. However, the style and mastery of language differed_from a learner to the other. For instance, five students wrote excellent essays. Their copies displayed a good mastery of content and linguistic knowledge with the exception of two or three spelling mistakes. They relied on their own style and their compositions were well organized. Fifteen students produced acceptable essays. Their compositions took the form of a combination of their own style as well as the content of the lectures but they made some spelling and grammar mistakes. The essays of twelve students revealed a low performance. Some of them had a difficulty in generating a correct response to the question while others had a difficulty in using language correctly. Eight students produced compositions whose level was below expectations because of their non mastery of language and their limited content knowledge.

In the module of psychopedagogy, the students had to answer two questions in two separate paragraphs. The first question was about the language derived from Latin and the second one was about the competency based approach. The majority of the students answered these questions in the form of separate titles or notes that provided definitions of the concepts involved in the questions.

In the module of literature, the examination question required the learners to analyse a passage taken from Joseph Conrad's <u>Heart of Darkness</u>. Most of the students started their essays by providing the biography of the writer and the

description of the characters. Then, they tried to give examples from the novel. This kind of essay organization led their answers to be incomplete as they wasted time in supplying unnecessary details.

In the module of TEFL, they were requested to explain the relationship between teaching and learning but some students focused on giving definitions of teaching, learning, approach, method, styles and strategies. Finally, they spoke about the teacher's role in few sentences. In this way their essays took the form of a summary of the lessons.

Therefore, the learners have problems in writing essays not only at the level of form but also at the level of the content of their pieces of writing. Nevertheless, a minority of students are able to produce good essays.

4.7- The Analysis of Students' Marks

The total number of the fourth year students is 153 learners divided into four groups. The average number of the students in each group was supposed to be 38 students.

This is why the sample of the learners' marks involved the grades of 38 students. The method used for the sample selection involved systematic sampling. Thus, the formula of selection was K= $\frac{153}{38}$ = 4,02 which means that every fourth member was selected from the target population. The grades of the students in the module of African civilization were utilized to make an analysis of these marks and to compare between the learners' performances.

The sample involved the marks of thirty eight (38) students. These scores were classified into two main categories. The first category was referred to as group D. It involved the students who did not get the average in the first and second examination and had to sit for the qualifying examination. The second category, called group E, included those students obtaining the average or more in the two examinations without needing to attend the qualifying one.

The following table gives an illustration of the marks of group D.

Students	Mark of the First Examination	Mark of the Second Examination	Average	Mark of the Qualifying Examination
S2	07	07	07	07
S4	06	07	06,5	
S5	08	10,5	09,25	
S6	08	08	08	11
S7	10,5	07	08,75	12
S8	07	09	08	10
S9	08	04	06	08
S12	03	05	04	07
S13	05	06	05,5	10
S14	06	10,5	08,25	11
S15	08	07	07,5	10
S16	09	06	07,5	09
S18	04	09	06,5	10
S19	06	08	07	
S22	06	12	09	
S23	07	06	06,5	11
S25	08	08	08	10
S26	10	08,5	09,25	
S28	08	06	07	11
S30	09	09	09	
S32	06	08	07	08
S34	06	10	08	
S35	09	08	08,5	
S36	03	06	04	05
S37	04	06	05	10
S38	05	08	06,5	10

Table 4.19 - The marks of group D

Therefore, 26 students out of (38), representing 68.42% out of the total group, got marks below the average in the two yearly examinations. The students of this subgroup can be classified into three categories. The first category includes those students obtaining the same mark in the first and the second examination or those getting approximate marks (the first mark plus or minus one point). This category involves nine (09) students representing 34,61% of the number of this group. They refer to those learners whose writing performance was somewhat static and did not change whether positively or negatively in the two examinations. The second category includes five (05) students whose marks decreased in the second examination. They represent 19,23% of the group. The third category involves twelve (12) students whose marks increased in the second examination. They represent 46,15% of the total number of this subgroup. The pie chart below gives an idea about the variation in the performance of the students obtaining marks below the average.
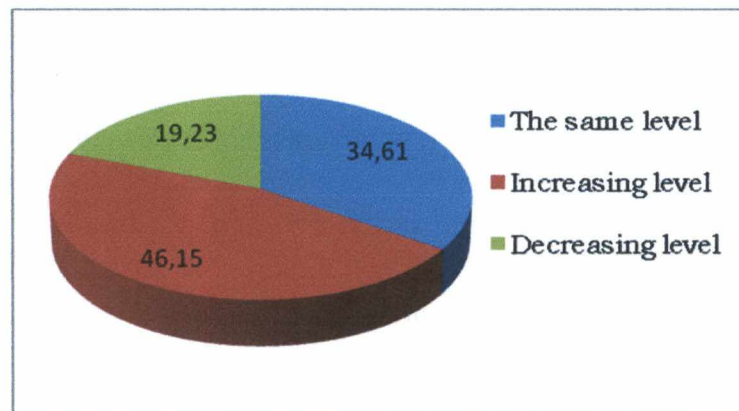
Pie chart 4.6 - The variation in the performance of group D
in the first and second examination

Thus, some students tend to keep the same level while the level of the majority of them improves in the second examination.

Regarding the performance of this group in the qualifying examination (see table 4.18), it should be mentioned that eight (08) students, representing 30,76% of the total number of this subgroup, did not attend such an examination. Concerning eighteen (18) students sitting for this examination, it may be stated that the majority

Students' numbers	Marks of the first examination	Marks of the second examination	Average
S1	14	14	14
S3	12,5	12,5	12,5
S10	12	09	10,5
S11	10,5	10,5	10,5
S17	11	12,5	11,75
S20	11	09	10
S21	10	10	10
S24	12	14	13
S27	10	10	10
S29	10	10	10
S31	14	13	13,5
S33	13,5	12,5	13

Table 4.20 - The marks of group E

Therefore, the students' marks range between ten and fourteen out of twenty. Eight students, representing 66,66% of the total number of this group, kept the same level. Two students, representing 16,66%, got lower marks in the second examination. Two other students, representing 16,66%, got better marks in the second examination.

This means that the level of the majority of the students getting the average in the two examinations tends to be static. Thus, their performance rarely varies from the first to the second examination.

As a conclusion, it can be said that twelve students out of (38), representing 31,57%, got the average in the first and the second examination. The same number obtained the average in the qualifying examination while fourteen (14) students out of (38), representing 36,84% did not get the average in the three examinations. These results are summarized in the following pie chart.
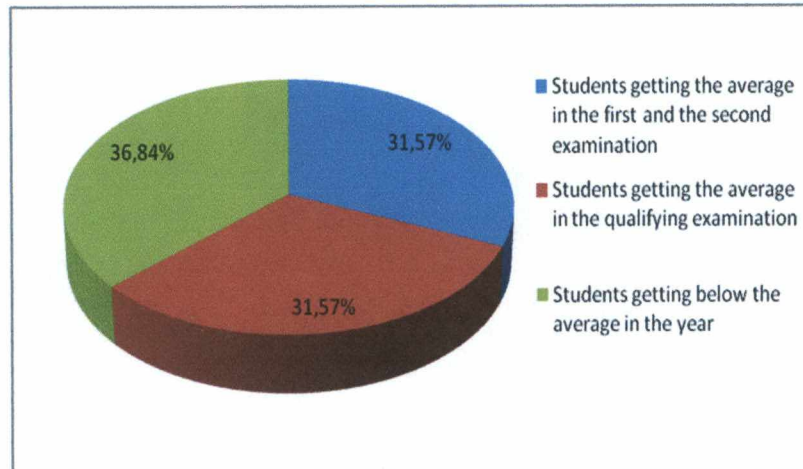
Pie chart 4.7 – The general performance of the studied sample

The general percentage of those students getting the average during the whole year is 63,14%. Thus, the majority of the students do not perform well in the first and the second examination. In this way, the qualifying examination gives them the chance to ameliorate their level of performance.

When comparing the results of the sample studied with those of all the fourth year students in the module of African civilization, it can be stated that there is a degree of resemblance in the learners' performance.

It should be mentioned that there are four groups in the fourth year. The number of students in group one is thirty five (35); group two includes thirty three (33) students; the number of learners in group three is forty one (41) while in group four, there are forty four (44) students. The following table provides information about the performance of the learners in each group.

Groups / Results	Group one		Group two		Group three		Group four	
	Number	Percentage	Number	percentage	Number	percentage	Number	percentage
Students getting the average in the first and second examination	15	42,85%	10	30,30%	19	46,34%	18	40,90%
Students getting the average in the qualifying examination	13	37,14%	11	33,33%	13	31,70%	08	18,18%
Students getting below the average in the year	07	20%	12	36,36%	09	21,95%	18	40,90%

Table 4.21 – The results of the fourth year students
in the module of African civilization

Therefore, the number of learners obtaining the average in the first and second examination in all the groups is sixty-two (62), representing 40,52% of the total number, which means that ninety-one (91) students got marks below the average in these examinations. Among the ninety-one students, nineteen (19) students representing 12,41% did not attend the qualifying examination while seventy-two (72) students, representing 47,05%, sat for this examination. Forty-five (45) students, representing 62,50% of the seventy-two learners got the average in the qualifying examination. They represent 29,41% of the total number of the fourth year students. On the other hand, twenty-seven (27) students, representing 37,50% of the seventy two learners, obtained marks below the average. They represent 17,64% of the total number of the fourth year students. The total number of the

learners who did not get the average in the whole year is forty-six (46) students, representing 30,06% of the fourth year students.

The following pie chart summarizes the results of the students in the module of African civilization.
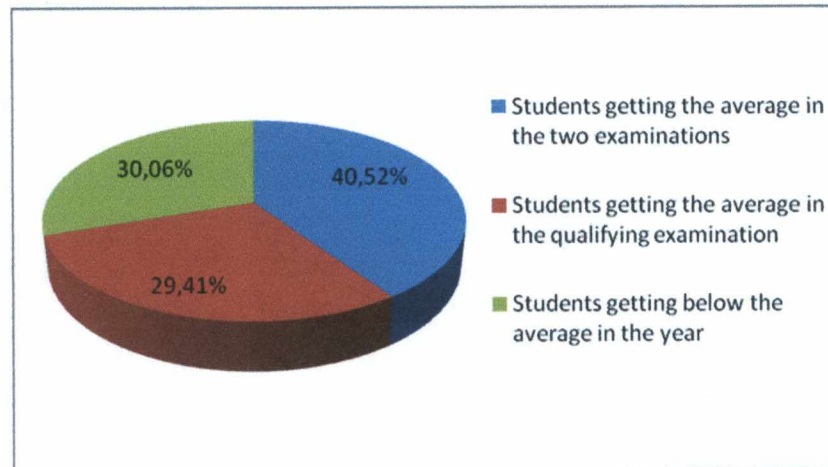
Pie chart 4.8 – The performance of the fourth year students
in the module of African Civilization

The general number of the fourth year students having the average in this module is 107 learners, representing 69,93%, while the general percentage of the learners obtaining the average among the sample studied is 63,14%. This means that the sample is approximately representative of the total population.

In addition to the study of the students' marks in a specific module, the comparison of the grades of a sample of five students was necessary to get information about the differences among the learners' performances. This small group of students included two girls (student A and B) and three boys (student C, D and E).

The following table provides an illustration about the grades of these learners.

Students	Student (A)		Student (B)		Student (C)		Student (D)		Student (E)	
Examinations / Modules	1st Exam	2nd Exam	1st Exam	2nd Exam	1st Exam	2nd Exam	1st Exam	2nd Exam	1st Exam	2nd Exam
Linguistics	06	09	09	11	10	11	08	10	08	10
TEFL	08	08	08	07	11,5	10	08,5	08	08	08
Psychopedagogy	12	13	11	13	13	15	13,5	10	11	13
Literature	11	11	07	08	12	10	09	08	08	08
African literature	13	11	08	09	08	10	10	12	08	08
African Civilization	09	11	08	09	08	09	05	10	08	08

Table 4.22 – A sample of students' marks in all the modules

From the above table, it can be noticed that the students performed differently in the various modules. Students (A) and (C) got the average in four modules out of six; students (B) and (D) obtained the average in two modules while student (E) got the average only in one module.

Concerning the module of linguistics, two students out of five, representing 40%, obtained the average. In TEFL, one student, representing 20%, performed well. In the module of psychopedagogy, all the learners obtained good marks. Two students, representing 40%, got the average in literature. The same number performed well in African literature. One student, representing 20%, obtained the average in African civilization. The following table gives the average marks of these students in all the six modules.

Students marks Modules	Student (A)	Student (B)	Student (C)	Student (D)	Student (E)
Linguistics	07,5	10	10,5	09	09
TEFL	08	07,5	10,75	08,25	08
Psychopedagogy	12,5	12	14	11,75	12
Literature	11	07,5	11	08,5	08
African literature	12	08,5	09	11	08
African civilization	10	08,5	08,5	07,5	08

Table 4.23 - Students' average grades

From the above table, it can be noticed that the performance of these students varies from a module to the other even if they seem to cope with the module of psycho pedagogy. Most of them did not perform well in the majority of the modules. This can be clearly explained by the mean of each module provided in the following table.

The Module	The Mean
Linguistics	9,2
TEFL	8,5
Psychopedagogy	12,45
Literature	9,2
Third world literature	9,7
African Civilization	8,5

Table 4.24 – The mean of each module

Therefore, the learners' degree of achievement may be affected by the nature of the taught module. The degree of performance of these students is illustrated in the following bar graph.
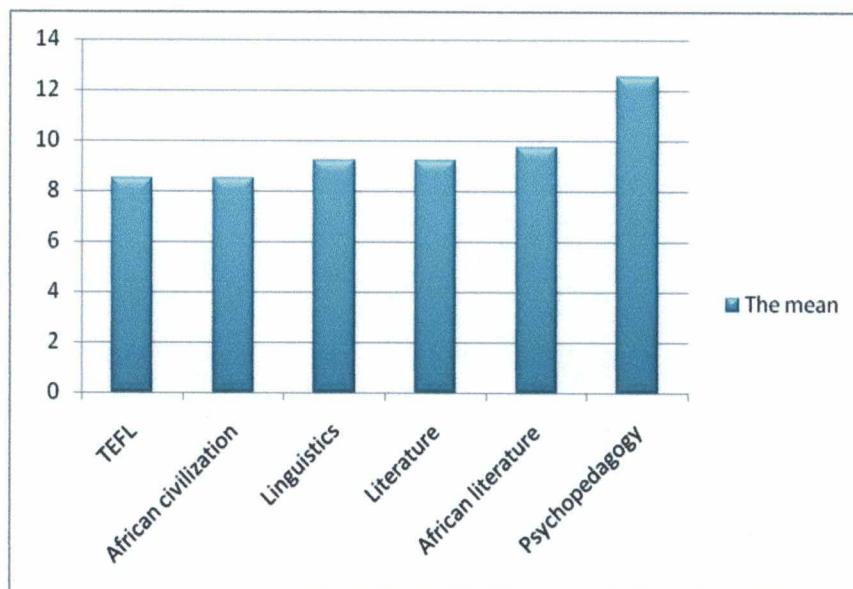
Bar graph 4.2 – The degree of the performance of the studied sample
in all the modules

Thus, the results of the sample of the five students reveal the impact of learners' individual differences on their performance.

In order to draw a clear picture of the degree of the learners' writing performance, it was necessary to analyse the examinations' results of all the fourth year students whose number was 153 learners.   The following table summarizes the performance of the fourth year students in all the modules.

Results \ Modules	Students getting the average in the first and second examination		Students getting the average in the qualifying examination	
	Number of students	Percentage	Number of students	Percentage
Literature	65	42,48%	17	11,11%
Psychopedagogy	44	28,75%	40	26,14%
Linguistics	85	55,55%	18	11,76%
TEFL	94	61,43%	24	15,68%
Third world literature	83	54,24%	28	18,30%
African civilization	62	40,52 %	45	29,41%

Table 4.25 – The performance of the fourth year students in all the modules

From the above table, it can be concluded that there is a degree of similarity between the performance of the learners in the first and second examination in the majority of the modules especially linguistics, TEFL and third world literature. In the qualifying examination, the results of the module of psychopedagogy resemble those of the module of African civilization while the results of the other modules do not greatly differ from each other.

Concerning the results of all the modules in the three examinations, it can be said that the degree of the performance of the learners is alike in the modules of literature and psychopedagogy. However, the results of these two modules slightly differ from those of linguistics, TEFL, third world literature and African civilization which display an approximate level of achievement. The table below gives an idea about the students' performance in all the modules throughout the whole year.

Modules / Results	Literature	Psycho-pedagogy	Linguistics	TEFL	Third world literature	African civilization
Number of students getting the average	82	84	103	118	111	107
Percentage	53,59%	54,90%	67,32%	77,12 %	72,54%	69,93 %

Table 4.26 - The learners' performance in all the modules during the whole year

Therefore, the performance of the fourth year students, in general, becomes acceptable only through the addition of the qualifying examination which gives the possibility to some learners to make a remedy to their deficiencies. Concerning the general results of the examinations of all the modules, it should be mentioned that seventy two (72) students, representing 47,05 % of the total number of learners, got the average in the first and second examination. In the qualifying examination,

twenty eight (28) students, representing 18,30 %, obtained the average. As a result, one hundred (100) students out of 153 learners performed well in the whole year. They represent 65,35% of the total population.

Thus, the majority of the learners, representing more than fifty percent of the total number of the fourth year students, find difficulties in the essay examinations of the first and second semester. Most of them reach the average in the qualifying examination. This means that some students perform well via a continuous process of assessment which leads them to practise writing essays within the testing situation to be able to achieve better results.

4.8- Data Interpretation

From the interviews, one can conclude that each teacher is in charge of a set of modules that are taught to distinct levels. The assessment of these modules is not the same since some of them are based on an analytic rating while others involve holistic rating. The table below gives a description of the types of scoring related to these modules.

Type of rating Modules	Analytic	Holistic
Oral expression		X
Grammar	X	
Written expression	X	X
Phonetics	X	
Reading	X	X
Linguistics		X
Literature		X
Civilization		X
Psychology		X
TEFL		X

Table 4.27 - The type of scoring the examinations of the taught modules

Thus, eight (08) modules out of ten (10) involve the employment of holistic scoring which means that 80% of the total number of the utilized examinations relies on holistic assessment. As far as the fourth year level is concerned, it can be stated that all the examinations are assessed through the use of holistic scoring as they require the learners to write essays. Even the students' dissertations are assessed holistically because the evaluation of the written projects and the oral presentation is based on the teachers' personal judgment.

Since the testing process is linked to teaching, one can conclude that assessment problems stem from learning difficulties which may be due to the learners' perception of the module, their way of receiving and elaborating the taught information as well as the first language interference in the learning process.

When sitting for examinations, the majority of the fourth year students have a feeling of anxiety which may have a great effect on the learners' results. The difficulties encountered by fourth year students in essay examinations are summarized in the following bar graph.
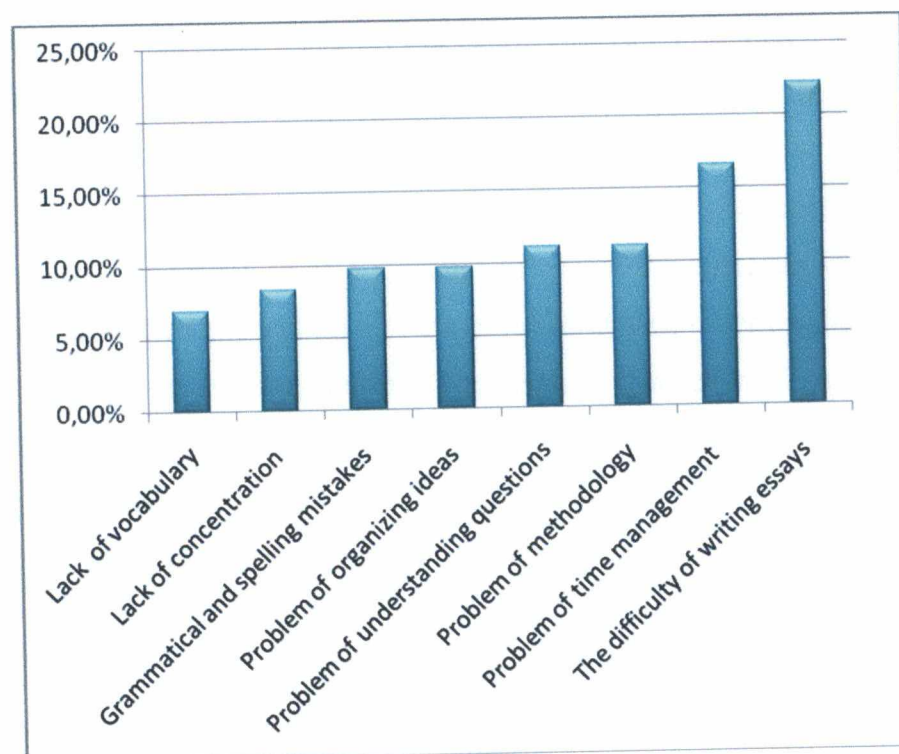
Bar graph 4.3 – The difficulties encountered by the students in examinations

The learners face certain problems in answering essay examinations because of their non-mastery of the language which is notably reflected through their deficiencies in the writing skill. Despite their weaknesses, the majority of the learners have a positive attitude towards essay examinations as they think that the latter enable them to practise writing and to demonstrate their degree of achievement. Nevertheless, not all of them are capable to provide a genuine personal work in the examinations because most of them produce essays which represent a sort of combination of the learners' understanding and the teachers' style denoted in the given lectures.

Even if it encompasses a sort of difficulty, holistic assessment remains a suitable testing method for the learners' linguistic and cognitive abilities. Thus, this procedure is most of the time utilized by the teachers of English. On the basis of table 4.7, one can infer that the questioned teachers are in charge of different modules focusing on teaching either language skills or a specific discipline.

Two teachers (teacher J and K) out of the twelve questioned teachers have to emphasize on the use of analytic scoring in the examinations of the modules they teach but they must resort to holistic scoring when assessing students' answers to the direct questions of the theoretical part of the examination of phonetics as well as the learners' paragraphs produced as a response to one of the exercises involved in the examination of written expression. On the other hand, two other teachers (teacher C and E) totally rely on holistic scoring. The remaining teachers focus more on the employment of holistic scoring in the modules they teach except for the ones dealing with reading, written expression and phonetics in which they use a combination of analytic and holistic rating.

Since testing and teaching are interrelated, the analysis of the teaching situation is important to get data about teachers' perception of their role. From the teachers' questionnaire, it can be stated that the teachers face certain difficulties

when teaching English such as the students' low proficiency and their lack of motivation. Concerning the purposes of teaching the settled modules, the majority of the teachers aim at making the students master the content knowledge of the taught module. Generally speaking, priority is given to language mastery in the modules related to language skills while more importance is directed towards content mastery in the ones linked to content areas. Regarding the questions of the examinations, they vary in their type according to the teaching objectives set up for each module. The fourth year students are tested through essay examinations.

The testing process creates certain burdens to the teachers who encounter difficulties in assessing the learners' abilities because of the different types of mistakes they produce. In addition to this, marking essays is a complex process involving the evaluation of different aspects of the learners' performance. The assessment of essays is based on rating the form and the content of the students' written products, but the teachers do not agree on the specific marks that should be awarded to form and content. Thus, some of them focus on grading the form while the others give more importance to the content of the essay. However, most of the teachers give equal grades to the two elements.

Even if the testing process is a difficult task, the teachers consider the construction of the questions of the examinations as an easy procedure as they extract them from the content of the lectures.

Concerning the types of mistakes produced by the students in essay examinations, the teachers agree on the fact that the learners display some weaknesses in the writing skill reflected in the lack of coherence and cohesion in their essays. However, they think that the general level of the students is average.

The main results of the interviews as well as the students' and teachers' questionnaires are summarized in the following table.

The main results of the interviews	The main results of the teachers' questionnaire	The main results of the students' questionnaire
- The predominance of holistic scoring - An exclusive use of holistic assessment at the fourth year level - Teachers' subjectivity	- The predominance of holistic scoring - The examination questions are based on the taught material - The difficulty of assessing essays - The issue of form versus content remains problematic	- Essay examinations as an opportunity to improve the writing skill - The difficulty of writing essays - Lack of genuine production - A feeling of anxiety created by the examinations

Table 4.28 – The summary of the results of the interviews and questionnaires

Regarding the results of the utilized tests, one can mention the difference in the teachers' way of referring to the learners' mistakes. The mean of the average marks of the tested students responding to the second test are illustrated in the following table.

Group A		Group B		Group C	
Students	The mean	Students	The average	Students	The average
P1	11	P6	07,66	P11	09
P2	10	P7	07,33	P12	09
P3	11,66	P8	06,66	P13	08
P4	10	P9	08,66	P14	07,33
P5	08	P10	11,66	P15	11,66

Table 4.29 – The results of the second test

When observing the average of the marks awarded by the teachers to the tested learners, one can state that four students out of five got the average in the first group. One student out of five obtained the average in the second group and the third group respectively. Therefore, the learners' performance differs from a student to the other.

The median of the first group is 10; the median of the second group is 07,66 while that of the third group is 09. Thus, the learners' performance may vary according to the learners' capacities and the teachers' way of assessing essays.

The median of the marks of all the fifteen students is 09 while the mean is 09,17 which means that the general level of the performance of these learners is approximate to the average.

concerning the variability among the assessors, it can be noticed that the difference in marking occurred at the level of the grades of two students (P2, P5) in the first group, two students (P6, P7) in the second group and two students (P14, P15) in the third group.

There was a distinction between the marks of two students out of five in each group, representing 20% of the total group. The general variability in the teachers' marking of the essays of the fifteen students is noticeable at the level of six students, representing 40% of the total number.

The following pie chart illustrates the degree of variability in the teachers' assessment of students' essays.
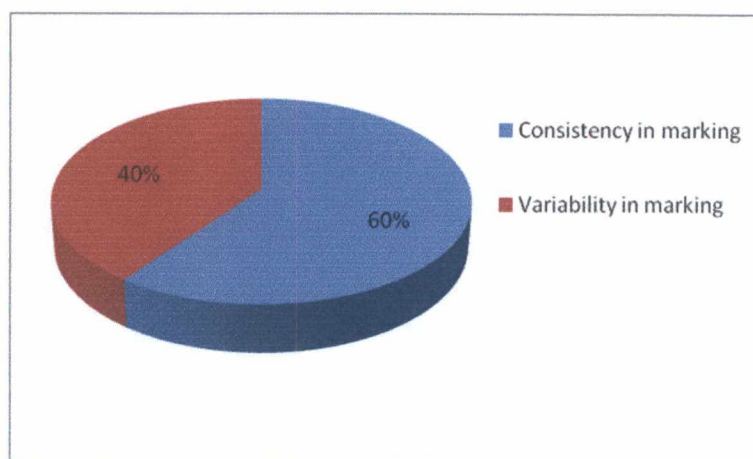
Pie chart 4.9 – The degree of variability in the teachers' scoring

The teachers perceive essay scoring approximately in the same way as the majority of the awarded marks reflect a certain agreement from the part of the teachers on the level of the learners' performance. The degree of variability in assessing is not very high.

As far as the employement of assessment as feedback is concerned, one can state that the majority of the teachers are aware of the role of testing in improving the learners' level through the comments made by the assessors about the students' essays as five teachers, representing 55,55% of the total number of the assessors contributing in the test's correction, correct the learners' mistakes or use notes to refer to them while the remaining teachers make a hint at those kinds of mistakes. Most of the tested students had difficulties in writing good essays as they produced a number of mistakes that affected the syntactic and semantic form of their writings. Also, they did not excel in demonstrating their knowledge of the discussed topic.

The analysis of the examination papers delineates a distinction in the teachers' way of identifying learners' mistakes and assessing essays. It also shows that the most common types of the mistakes made by the students are spelling and grammatical mistakes in addition to the generation of some ambiguous sentences. Concerning the learners' method of organizing and structuring the content of their answers, it may be stated that the studens' essays often happen to be a slightly modified reproduction of the lectures.

The analysis of the students' marks in the module of African civilization gives an idea about the learners' performance. The mean of the marks of the sample in the two examinations is 08,58 while the median is 08,12. When replacing the marks of the students who did not get the average with their marks in the qualifying examination, the mean becomes 09,87 and the median is 10. Thus the students' performance in the two examinations is not very satisfactory but it reaches the average thanks to the results of the qualifying examination. The careful examination of the grades of group D which refers to the students getting marks below the

average reveals that the modes are (06), (05), (07) and (08) which means that the learners' low marks often range between 06 and 08/20.

On the other hand, the mode of group E is 10 which means that most of the students getting the average are awarded the mark of 10/20. Generally speaking the learners' performance may be improved through the use of successive examinations which can be considered as a sort of additional practice and undirational process. The mean of the marks of the fourth year students in the module of African civilization in the first and the second examination is 08,72. The general mean obtained after the subtitution of the low marks of the students with the ones they obtained in the qualifying examination is 10,08. This means that there is an improvement in the learners' performance by the end of the academic year.

The study of the marks of a sample of students reveals certain differences in the performance of each learner. These distinctions involve an alternation in the learners' results from one examination to the other as well as a difference in the results obtained from one module to another. In addition to this, the students' results are distinct from each other. Generally speaking, the majority of the students do not perform well in the first and the second examination.

Regarding the results of the fourth year students in all the modules in the first and second examination, one can notice a distinction in the results obtained from a module to another. These results are illustrated in the following bar graph.
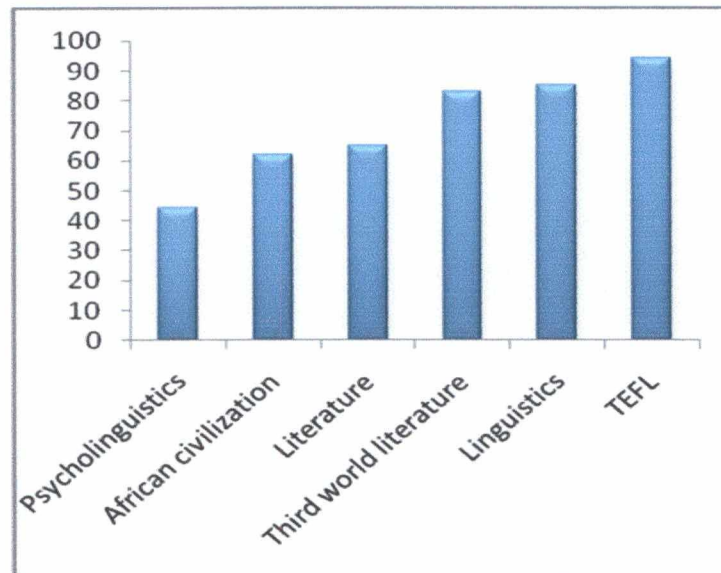
Bar graph 4.4 – The students' performance in the first and second examinations

The learners' performance in the two examinations is approximately similar in the modules of TEFL, linguistics and third world literature whose results slightly differ from those of the modules of literature and African civilization which are nearly identical in the number of students performing well in these modules. The point of divergence concerns the module of psychopedagogy which reflects a great degree of distinction from the results of the other modules. However, the differences existing among the modules regarding the total number of students getting the average in the whole year seem likely to diminish thanks to the results of the qualifying examination. The degree of the learners' performance in the academic year is elicited in the bargraph below.
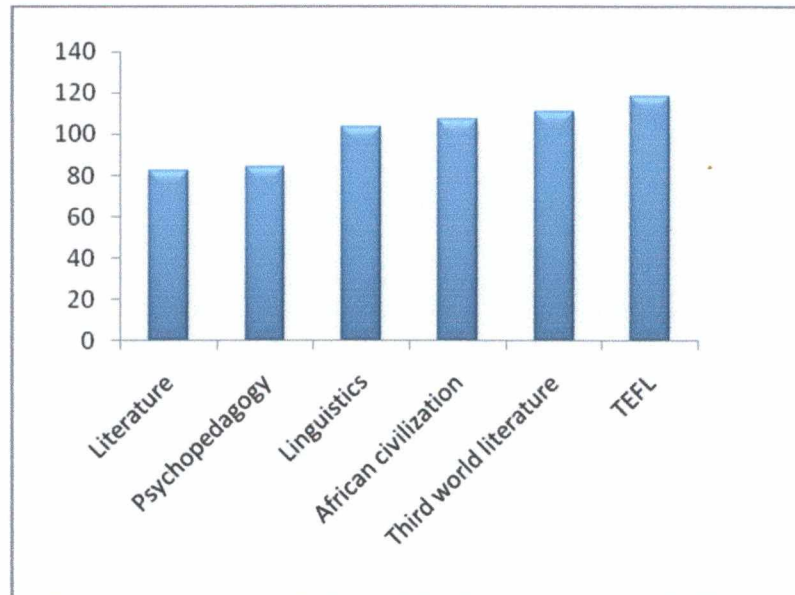
Bar graph 4.5 – The learners' degree of performace in all the modules.

Therefore, the general performance of the fourth year students in all the modules is fairly good since there is an increase in the total number of students obtaining average marks. As a result, the learners' ability to write essays may change from a situation to another. An improvement may occur at the level of students' written production through time after the utilization of a continuous form of assessment. The following pie chart gives an idea about the degree of achievement of the students in essay examinations.
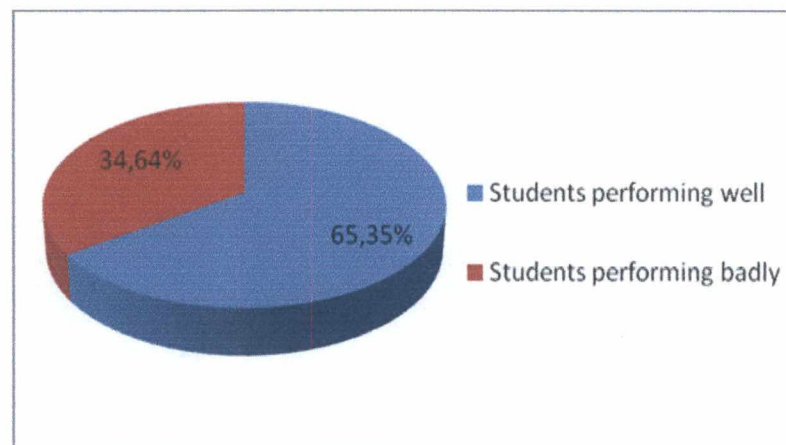
Pie chart 4.10 – The general performance of the fourth year students

Thus, the majority of the fourth year students manage to reach an acceptable level in writing essays by the end of the year.

4.9- Discussion

This research work aimed at providing information about the employment of holistic scoring for the assessment of students' compositions as well as the kinds of difficulties related to this process. Thus, two main issues were taken into consideration including the learners' writing performance and the teachers' scoring of essays. The analysis and interpretation of the gathered data constitute a basis for the formulation of the following conclusions which concern the predominance of holistic rating and the problems related to such a procedure as well as the validation of holistic scoring.

4.9.1- The Predominance of Holistic Scoring

The testing situation at higher education in Algeria is characterized by the employment of holistic scoring as a major tool for assessing writing since most of the modules taught at the English section are tested holistically. Although the examinations of some modules like phonetics, reading and grammar are based on analytic scoring, they often involve a set of exercises that oblige the teachers to rely on their appreciation of the given answer which means that a combination of holistic and analytic forms of rating is necessary in such situations. On the other hand, the majority of examinations involve essay writing. As a result, the testing process is based on the utilization of holistic methods characterized mainly by the teachers' personal judgement of the written products.

The use of holistic scoring for the assessment of students essays at university level is considered as a means to rate the learners' writing performance and mastery of the target language. The main purpose of essay examinations is to allow the

learners to display their knowledge, their level of understanding, their composition skills and degree of mastery of the different elements of the target language.

Thus, they become relevant tools of assessment at higher level of education. The most appropriate method for rating such examinations is holistic scoring which gives the possibility to capture all the features of writing and provide a general evaluation of the learner's performance as a complete entity.

The fact that this procedure is not time consuming helps the teachers to correct the large number of students' copies within the limited period of time designed for giving back the marks especially by the end of the year.

In fact, research has proved that essay examinations are the most suitable direct methods of assessing students' writing performance. Van Blerkom (2008) asserts that essay examinations are quite popular. They are the most common tools of assessment at the level of higher education since they give the possibility to measure higher-level cognitive skills like application, analysis and evaluation. Also, essay questions help the learners to show their competence concerning the subject matter and their capacity to engage in critical thinking. This is why they are frequently used for testing the ability of older students to express their ideas in a written form. According to Tynjälä (2001), essay examinations enable the learners to make different kinds of knowledge transformations such as reasoning, questioning, metacommenting, using schemata and restating since the process of writing compositions deals with discourse synthesis.

Holistic rating is employed for scoring essays as the assessors have to provide a holistic score for the students' responses by judging the written product as a single entity.

4.9.2 - Problems of Holistic Scoring

There are certain problems that are often associated with holistic scoring such as subjectivity, unreliability and the issue of form versus content.

4.9.2.1- Subjectivity

Holistic scoring is subjective as it involves the teachers' personal judgement. The interviews revealed that subjectivity appears at two main levels that include the examinations and dissertation assessment. Some teachers may be diverted from the objective of testing by the student's style or good command of language which leads them to ignore the learner's incapacity to achieve critical thinking and good analysis of the topic of the essay examination.

Here, one can cite Elander (2003) who claims that research conducted about the scorers' bias has revealed that the marking process may be influenced by certain factors called the contrast effects which refer to the impact of the quality of the work assessed beforehand. Thus, good work can be scored more favourably when preceded by essays of poor quality and poor work can be severely judged when following essays of good quality.

Another aspect of teachers' subjectivity concerns the assessment of dissertations. The interviewed teachers affirmed that they highly evaluated the work of the students they supervised. They added that they found difficulties in rating the learners' memoir because of bias. To consolidate this statement, it is essential to mention that some studies focused on the effect of bias on the marking of students' projects. For instance, research revealed that the scoring of projects could be influenced by the personal knowledge of the learner or the fact that the scorer is the supervisor of the student. Also, research indicated that supervisors could be more biased than examiners when marking the learners' work.

In fact, subjectivity always exists in the process of scoring as it is a human feature. It remains a factor that interferes in the assessment of essays.

Lynch (2003) believes that rating is unavoidably subjective. Subjectivity can be controlled via raters' training or the use of specific scoring criteria. However, it can not be totally eliminated since the scoring procedure relies on the scorer's judgement even when assessment criteria are employed.

Bias influences the assessors' way of marking which leads to the variability of the scores. This means that subjectivity can affect the reliability of scoring.

4.9.2.2 - Unreliability

The use of tests as research instruments aimed at providing information about scorers' reliability taking into consideration both interrater reliability and intrarater reliability. The result was that the degree of variability between the scores tended to increase in certain cases while it diminished in other situations. There was a sort of consistency in marking between certain teachers while a few of them deviated in the scoring of some copies. Thus, the scores were not totally unreliable but they can not be considered as completely consistent.

One can infer that the extent of the scorers' reliability is acceptable although it is not very high. This can be justified by the fact that reliability is often totally achieved when scoring tests of physical sciences or when using objective language tests. On the other hand, in the domain of social sciences, it is difficult to measure and to increase reliability. This rule is also valid when testing language via direct methods of assessment since the learners' differences as well as the scorers' differences make it impossible to agree on one definite score to each individual. In this context, Van Blerkom (2008) affirms that the typical tests that measure learners' knowledge or ability are about 70% reliable.

Heywood (2000) notes that the first study about scorers reliability took place in the United States in 1908 through the research of Max Meyer who concluded that the marks of examiners were highly inconsistent. This issue was discussed in Britain in 1935 thanks to the work of Phillip Joseph Hartog and Edmund Cecil Rhodes who analysed scores provided by various examiners and found that these marks were variable. In 1957, Paul Diederich asked 53 experts to score 300 short essays and sort them out into nine (9) categories of merit. The result of this experiment was that 34% of the essays received all the nine grades and no essay obtained less than five grades out of nine (Rowntree, 1987).

Elander (2003) reports that similar studies were made by Sara Weigle in 1998, as well as Greatorex Baird and J.F. Bell in 2002. The result of these studies confirmed that there was not a total agreement between the scorers. Heywood (2000) states that research has proved that the same paper can be given different marks whether it is scored by different teachers or by the same teacher who rates it several times.

It can be assumed that the debate over the issue of scorers' reliability will not stop as the variability in the scores awarded by different assessors is a recurrent aspect of the direct assessment of writing.

Interrater reliability is often affected by scorers' experiences, their personal preferences and their expectations from the students' writing. Rowntree (1987) asserts that interrater inconsistencies can never be entirely resolved due to the differences in the assessors' judgement and their distinct orientations toward the subject. Some scorers may give more importance to a certain quality of the student's performance while other assessors reading the same paper may not consider this element as very important and will focus on other aspects of the learner's writing. Intrarater reliability is affected by the conditions or timing of the various rating occasions and the scorer's personal judgement.

Therefore, different factors interfere in the scoring procedure so that it is impossible to ascertain a very high degree of consistency. *"Writing tests cannot be 100%, and are rarely more than 80% reliable"* (Hamp-Lyons, 2003: 163). This is due to the fact that the assessors are human beings. Thus, they have preferences for certain ideas; they may dislike some arguments and they can change their opinion from a day to the other. Hamp-Lyons (2003) claims that writing assessment in the United States and other countries is now considered as 75% reliable. The expectation of reliability when scoring essays is problematic mainly because writing is a complex skill that involves a combination of various processes and language elements.

Hyland (2003) states that indirect measures were employed to avoid the problems of reliability. However, researchers claim that indirect methods are reliable but they are not valid since they are concerned with accuracy rather than communication. Consequently, it is asserted that despite the fact that holistic scoring is not totally reliable, it possesses strong validity as it focuses on the learners' performance and their ability to use the language for interactive and communicative purposes.

4.9.2.3 - Form versus Content

The teachers' questionnaire revealed that there was not a total agreement about whether to give more importance to the form of the scored essay or its content. The majority of the teachers take into consideration the form and the content of the students' essays; some of them slightly favour the former over the latter; a few teachers focus widely on form more than content while a minority ignores form and puts emphasis just on content.

Research has dealt with this issue but no clear answer has been provided. Fathman and Whalley (1990) assert that some studies which focused on how teachers should respond to students' writing revealed that many teachers were interested in grammatical correctness. However, the main question concerned where teachers should focus their attention. Thus, disagreement exists over the issues of form and content.

Sakyi (2000) affirms that some research studies revealed that teachers gave importance to the use of language especially the absence of errors in the learners' compositions while other studies proved that some teachers focused more on the content of the essay. This means that more research is needed concerning this topic. The form of the essay which refers to language use including the knowledge of grammar and vocabulary is an essential element that demonstrates the students' mastery of the target language. The content of the composition which is represented through ideas and their organization is also important as it reflects the learners' knowledge of the topic or subject-matter. The preference of form over content or

vice versa depends on the teachers' perception of the purposes of writing assessment and their expectations of the learners' performance.

4.9.2.4 – The Variables Related to Students' Writing Performance

The students' writing performance interacts with various variables that may have a positive or negative impact on its quality concerning the independent variable which is linked to the input, it can be stated that the students are supplied with various kinds of information that enable them to cope with the testing situation. The data collected from the interviews revealed that the program of the fourth year focuses on teaching the learners a set of modules that deal with different disciplines or content areas for the sake of developing the students' content knowledge. Therefore, the taught information develop the learners' knowledge about the different topics that can be involved in the examinations.

Regarding the intervening variable which is related to the students' exposure to the written form of the target language, one can confirm that the learners are taught about the writing skill during the three years of their study at university in the module of written expression. In addition to this, the learners are exposed to the written form in the different modules as they have to take notes during the lectures. Throughout the three years of their graduation studies, they have to use this skill in the majority of the examinations. In the fourth year, all the examinations are in a written form.

The learners' exposure to the writing skill is a very important aspect in the assessment of this skill as it influences the learners' performance. Harrington et al. (2003) state that research revealed that there was a relationship between knowledge in the subject area, the writing skill and the ability to transfer this knowledge in a communicative situation. This means that the teaching program and the extent of the students' exposure to the writing skill have a great impact on their performance.

Another aspect that may affect the learners' performance is the topic which is referred to as the hidden variable. From the results of the test, it can be inferred that the topic may influence the learners according to their level and their

personality. The students may perform well in certain modules just because they like the topic or they possess a considerable amount of knowledge about the given subject. Concerning the examination questions, one can say that the teachers either opt for direct questions or quotations whose content is related to the studied subjects. There is no problem with the wording of the questions which are often understood by the majority of the learners. However, it is difficult to provide evidence about the impact of the topic on the learners' performance as certain mental and psychological factors may interfere in this process. Ruth and Murphy (1988) assert that some researchers acknowledge the effect of question wording on the learners' responses. There were some studies which focused on the impact of the topic on the students' writing performance. For instance, in 1953, Gerald Kincaid made a study about this issue and concluded that the assigned topic affected the performance of poor writers while it did not have an impact on the performance of good writers. In 1963, Richard Braddock, Richard Lloyd Jones and Lowell Schoer produced a reinterpretation of past studies which led to the conclusion that the topic effect has to be controlled although it is not easy to do so. In 1982, James Hoetker conducted a study which enabled him to affirm that there was not enough information about the nature of writing tasks and their effects. He added that research has often ignored three main issues related to essay examinations which include validity, the topics and the writer.

Therefore, the wording of the essay question and the nature of the topic may lead the learners to respond in a specific way according to their own interpretation of the given question. The writing task can function as a stimulus or a constraint depending on how the question is formulated and the extent to which the students manage to discuss it. However, this effect may vary from a learner to the other according to the student's knowledge of the topic, his appreciation and interpretation of this subject as well as his mastery of the language. It is difficult to understand the real effect of the topic on the learners' performance. Even the procedures of item analysis cannot be employed to find out how essay questions will function in a real testing situation.

Concerning the dependent variable which is the learners' writing performance, one can state that the students differ in their degree of achievement. From the data collected through various instruments in this research work, one can conclude that the writing performance varied from a learner to another. At the same time, the achievement of each learner may differ from an examination to the other. This means that the students performance is not only affected by the different variables discussed earlier but it is also influenced by some factors that may be sometimes difficult to examine.

Rupp and Köller (2008) state that research has focused on the relationship between the writing ability and the individual factors which include memory capacity, cognitive processes, motivation and affect. They also assert that the results of some studies highlight the necessity of acquiring linguistic, intercultural and methodological competencies as well as communication strategies in order to become successful writers. The linguistic competencies refer to the successful application of linguistic resources especially the knowledge of vocabulary and grammar. The intercultural competencies are related to the students' knowledge of the rules of communication in addition to the perceptions and prejudices of their own culture and the foreign culture. The methodological competencies mean the application of methods and strategies in order to produce written texts. The communication strategies refer to the application of the meta-cognitive principles of planning, execution and monitoring.

Weigle (2002) mentions that researchers point out that the quality of students' writing may be influenced by motivational factors as well as affective variables like writing anxiety or writing apprehension. Tynjälä (2001) claims that research has shown that the learners employ different learning strategies. These strategies may involve a deep approach that concentrates on meaning and understanding or a surface approach that focuses on memorizing and reproduction. Tynjälä (2001) also notes that some research studies have revealed that certain students often employ a deep approach when responding to essay examinations while others use a surface approach.

Different factors may intervene in the process of essay writing leading the students to produce distinct compositions according to their understanding, their motivation and the strategies they utilize to answer the essay question. Therefore, it is difficult to explain the differences in the learners' achievement as well as the variability in the students' scores from an examination to the other.

4.9.2.5 - The Validation of Holistic Scoring

The data collected from the teachers' questionnaire confirmed that the main objectives of teaching the various modules included the learners' mastery of the English language as well as their mastery of the content knowledge related to the studied discipline. On the other hand, few teachers viewed the mastery of language as the chief purpose while others regarded the mastery of the content as the main objective of teaching. The teachers' perception of these objectives was reflected in their opinions about whether to correct the form or content of the students' essays.

Therefore, the learners' compositions are scored according to the assessment purposes which are linked to the teaching objectives. In fact, the utilization of holistic scoring for assessing the students' writing performance aims at testing the learners' ability to use the language to convey specific information. The assessment results are employed for certification purposes. This means that the learners' achievement is considered as an indicator of the extent of their mastery of the taught modules. The obtained scores determine whether they pass or fail. In the case of the fourth year students, the examination results provide a basis for making decisions concerning whether the learners can get a licence degree or not. As a conclusion, one can state that holistic scoring is somewhat valid as it fulfils the purpose it is intended to achieve. However, the degree of the validity of such a procedure cannot be predicted as various factors may affect the students' scores and lead the scoring consistency to fluctuate.

The quest for the validity of holistic scoring relies on the process of the validation of this instrument via the analysis of the collected data. In this research

work, the gathered information suggest that the majority of the teachers quite agree on the measurement of the students' performance. Despite the degree of unreliability that may exist in the scoring of some assessors, there is an agreement on the kinds of decisions that will result from the analysis of the obtained scores. However, the question that is often raised is the possibility of considering holistic scoring as being more valid despite the existence of a certain inconsistency in scoring. This issue has been tackled by researchers.

Several studies focused on the validation of holistic scoring. Sakyi (2000) claims that William McColly was the first to make research about the validity of holistically scored compositions. Such concerns are still raised by researchers. McArthur et al. (2006) insist on the importance of making a distinction between reliability and validity since the latter stems from the need to make the best decisions for the students. A study made by Sakyi (2000) revealed that the holistic scores of written compositions were affected by different factors. The conclusion was that holistic scoring could become reliable through training or when raters had similar backgrounds but there must be a certain agreement among the scorers about the decisions that should be made from the assessment results.

Urquhart and Mciver (2005) assert that most researchers support the employment of holistic scoring for assessing writing. Elbow (2000) states that research focuses on scorers' consistency so that there is a tension between validity and reliability. He adds that good agreement between assessors is not a proof of validity which is linked to what the teachers are trying to look at. Elbow (2000) suggests that the way assessors actually read essays is a mirror of validity so that it makes more sense to focus on validity and allow reliability to suffer. In addition to this, Lynch (2003) asserts that disagreement between the scorers should be seen as a source of validity rather than an indication of unreliability since different scores provide a variety of information about the learners' performance.

As a conclusion, one can say that the validity of holistic scoring relies on the quality of the data provided by the scores which will establish a basis for making decisions about the students. Since differences in scoring will always exist, it is

better to give more importance to the achievement of the purposes of assessment via the accurate interpretation of the examination results.

4.10 – Conclusion

This chapter has focused on the interpretation of the collected data. Thus, some explanations were provided about the difficulties faced by the students when sitting for essay examinations as well as the teachers' perceptions of the teaching and testing purposes. This chapter also involved a brief discussion of the main results of the present case study. The conclusion was that holistic scoring could be more valid than reliable. However, the teachers have to put the emphasis on the information elicited from the assessment results for the sake of making decisions about the learners. Holistic scoring can raise certain problems, but it may become more effective through various means. This is why the next chapter will be devoted to the suggestion of some techniques and strategies that may develop this procedure and improve the process of writing assessment.

CHAPTER FIVE:

SUGGESTIONS AND RECOMMENDATIONS

CHAPTER FIVE:

SUGGESTIONS AND RECOMMENDATIONS

5.1- Introduction

This chapter attempts to provide some suggestions and recommendations that can contribute in improving the teaching and assessment of writing. First, it puts the emphasis on the necessity of increasing the efficiency of essay examinations as well as maintaining and developing holistic scoring. Next, the chapter discusses the idea of employing analytic marking scales for scoring compositions and the construction of scoring rubrics so that to facilitate the assessors' task. Then, some strategies for assessing dissertations as well as the provision of feedback through assessment, the application of distributive evaluation and the implementation of product analysis are discussed. Also, the suggestion of using the six traits for teaching and assessing writing is proposed. Finally, the chapter sheds light on the fact that the development of the testing of writing is closely related to the improvement of the instruction of this skill which can take place by utilizing assessment as means of teaching writing and raising students' motivation in addition to the use of various techniques for teaching this skill.

5.2- Increasing the Efficiency of Essay Examinations

Essay examinations are seen as very essential tools for the evaluation of the learners' writing performance. They enable the learners to demonstrate their composition skills. *"Essay items have been around for so long that it is a good bet Socrates once asked Plato to describe Greek-style democracy in an essay of at least 500 words"* (Popham, 2003: 92).

Thus, compositions enable the learners to express themselves through writing their ideas in a detailed manner. They exhibit the students' competence in discussing the given topic. Essays are considered as the most popular method of measuring the learners' ability to engage in critical thinking. They are useful instruments of assessing higher-level cognitive skills like analysis, synthesis, application and evaluation (Van Blerkom, 2008).

Research has revealed that the utilization of essay examinations for the assessment of writing urges the learners to study more efficiently as they have to put the emphasis on broad issues and interrelationships between concepts. This leads the test results to provide a better evaluation of the students' performance. Also, essay examinations give the teachers the possibility to judge the quality of the learners' thinking and the depth of their understanding as well as their progress and the difficulties they are encountering (Davies, 2009).

These types of examinations have various potentials. They are used to test the learners' understanding, knowledge and ability to recall information in addition to their written communication skill. They possess a degree of reliability as different assessors are likely to award the same mark to the same paper. They are also characterized by robustness which refers to their resistance to abuse by students like cheating and plagiarism (Nation, 2008).

In order to make the essay examination effective, several aspects of the testing situation have to be taken into consideration. Hamp-Lyons (1990) asserts that the main elements that affect the process of assessment are: the task, the writer, the reader and the scoring procedure. The writing task refers to the topic or the question designed to assess the learners' performance. Its effectiveness depends on the prompt attributes and the period of time allotted to the examination. The prompt attributes mean the characteristics of the task. They involve the components that formulate what the teachers have to do.

> *Prompt attributes for writing include topic, content for the prompt, purpose, audience, cultural expectations and rhetorical functions. It should also be noted that the choice of the topic can limit or constrain the types of language that will be produced by the test takers*   (Lynch, 2003: 51-52).

This means that the nature of the subject that has to be discussed by the learners in their essays is very important as it has an impact on their performance. The students' responses may vary according to the given topic. The prompt attributes focus on the task's content and the time limits within which the learners have to answer the question. They are closely linked to the response attributes which mean how the testees will interact with the prompt. Writing tasks should give the tester the possibility to make appropriate inferences about the students' ability to display their knowledge and linguistic competence through writing (Weigle, 2002).

The design of a writing task is based on four elements which include the rubric, the prompt, the expected response and the post-task evaluation. The rubric refers to the instructions that clarify to the test takers how to carry out the task. It may involve the description of the purpose of assessment and the procedure of responding like explaining if the learners have to produce sentences, a paragraph or an essay in addition to the evaluation criteria which can be expressed in a statement that mentions the feature that is given more importance. The second component of a task is the prompt which means the stimulus to which the learner has to respond. It is linked to the topic that has to be discussed. Another important element that is related to the development of the test item is the expected response which concerns how the learners are intended to perform. The post-task evaluation is also necessary as it implies assessing the effectiveness of the task (Hyland, 2003).

When constructing essay items, the teacher has to take into account various features. The task should be circumscribed by explicitly expressing its content in order not to leave any doubt to the learners about the expected answer. Second, the point value of the question has to be specified to inform the learners about the test's valuation. For instance, the teacher can mention that the score of the essay is twenty (20) points. In addition to this, an acceptable response length should be recommended. For example, the students can be told that their composition should

not exceed 200 or 300 words. Finally, the quality of the essay question has to be measured by producing a trial response to the designed topic (Popham, 2003).

Essay examinations are easy to construct because they do not take much time in their construction. However, the teacher has to provide clear questions that will not be seen by the students as a trick. Jacobs (2004) argues that clear essay questions are easy to answer by the students and easy to score by the teacher. This is why the words used for the essay should be carefully chosen. For instance, asking the learners to discuss a given of view may be a little bit vague. Also, it may require specific knowledge from the part of the students to make a sort of discussion. However, not all the learners will be able to do so. Therefore, it is preferable to use words like 'comment', 'explain' or 'compare'.

The use of essay examinations provides a realistic situation to the students since they have to express their ideas and thought. On the contrary to objective tests including true-false or multiple choice items which measure the students' ability to recall what has seen to them as well as their reasoning and comprehension, essay examinations are employed to test complex cognitive processes involving the students' way of thinking, organization of ideas and ability to communicate using the written form of a language. Thus, essay questions should focus on demonstrating the thinking processes rather than factual knowledge.

According to Jacobs (2004), the essay question has to represent a problem situation that tests the students' ability to use knowledge in order to analyse, justify and explain. He suggests that the teacher may use essay questions that test specific cognitive skills using the different levels of the taxonomy of teaching objectives (knowledge, comprehension, application, analysis, synthesis and evaluation) listed by Benjamin Bloom (1956).

The following table provides an illustration of the different types of essay questions that may be employed to assess different kinds of thinking abilities.

Bloom's cognitive levels	Examples of essay questions	Assessment objectives
Knowledge	- Identify the different teaching methods. - Define cognitive strategies.	Testing the students' ability to remember factors or concepts and define them
Comprehension	Explain the following statement: 'Teaching is an art'	Assessing the students' ability to explain and summarize what has been taught to them.
Application	Use the different language properties to explain the difference between human and animal communication	Testing the learners' ability to make use of the taught material to clarify a certain concept or idea
Analysis	- Compare the style of Jane Austen and Charles Dickens - Why do linguists consider speech as a primary form of language?	Assessing the students' ability to make a distinction between various concepts by making comparison or providing arguments.
Synthesis	- Write a short story using the ideas provided in the given poem. - Imagine a different ending to the story of Okonkwo in *Things Falls Apart*	Testing the learners' capacity to develop ideas and produce creative pieces of work on the basis of the taught information
Evaluation	- Evaluate the degree of effectiveness of the competency based approach. - Criticize the principles of generative grammar.	Assessing the students' ability to make a judgment or criticize certain ideas on the basis of the taught material

Table 5.1 – Essay questions according to the different cognitive levels.

Another aspect that influences the writing test is the writer who refers to the test taker. In fact, the testees may respond differently to the given essay examination because of their distinct characteristics which involve their cultural background, gender, native language and personality.

These features have to be taken into consideration during the scoring process as the testees actual performance may diverge from the teachers' expectations. The affective and moral aspects of the students can have a great effect on their written production. Therefore, the essays produced by the learners are often affected by their personal opinion which may produce a sense of subjectivity within their compositions. Training the students to use logical thinking in their writing will help them to limit the scope of divergence between the learners' responses and the teachers' expectations.

If the essays are written according to the writers' reactions, they are also scored according to the scorers' personal judgement. This means that the reader who refers to the rater plays an important role in the assessment of writing. The readers score the learners' composition on the basis of a set of criteria determining the correctness of the response. Their way of interacting with the learners' essays has a tremendous impact on the scoring results. In order to decrease the scorers' subjectivity, researchers suggest that reader or rater training can lead the assessors to read the students' essays consistently. Also, the use of rubrics can help the readers to focus on certain specific criteria rather than making a vague judgement (Hyland, 2003).

The other aspect that is considered as a critical element in the process of assessment is the scoring procedure which enables the reader to interpret the writer's ideas. The assessor has to decide about the rating procedure that should be employed for assessing the learners' compositions. This method may take the form of an analytic, holistic or trait based scoring. However, holistic scoring can be a suitable technique for rating essays on the condition that the scorers are cautiously trained to respond to the learners' written product.

Holistic reading is based on the view that there are inherent qualities of written text which are greater than the sum of the text's countable elements, and that this quality can be recognized only by carefully selected and trained readers, not by any objectifiable means. (Hamp-Lyons, 1990: 79)

The scoring procedure guides the teacher in his way of perceiving the piece of writing. It relies on the use of a rating scale that provides descriptions of the expected performance so that the rater can judge the learners' compositions through the use of explicit features of writing that determine the level of achievement that is measured.

The final result of the scoring process is formulated in the form of a score which is considered as an interpretation of the students' writing ability that can be utilized for making decisions about the learners. Thus, *"...a score in a writing assessment is the outcome of an interaction that involves not merely the test taker and the test, but the test taker, the prompt or task, the written text itself, the rater(s) and the rating scale"* (Weigle, 2002: 108). This means that different elements contribute in the testing process. The essay examination depends on various components. The test taker responds to the task according to his perception of the topic, his linguistic knowledge and culture. He produces the written text which is called the 'script' in Britain and Australia but in the United States of America, it is often referred to as essay or composition while the term script is not widely employed.

The written text is the learners' response to the task which is considered as a medium between the testee and the reader who relates the produced essay to the features of writing via a rating scale that facilitates the scorer's understanding of the criteria that qualify a good composition. *"Each aspect-task, writer, scoring procedure and reader-interacts with the others creating a complex network of effects which to date has eluded our efforts to control"* (Hamp-Lyons, 1990: 82).

Essay examinations are affected by various aspects that lead the assessment results to be unpredictable as each element may intervene in a positive or negative manner. As a result, the actual performance of the learners may not reflect their real writing level. Hence, all these aspects have to be controlled in order to ensure the efficiency of the testing methods.

The improvement of the quality of essay examinations depends on the adjustment of the testing situation to the purposes of assessment. It relies on the development of essay prompts that test the learners' linguistic and cognitive abilities. The use of various items from which the testees have to select one topic may enable the learners to express themselves freely (Welch, 2006). This variety of topics avoids penalising or favouring some learners because of special background knowledge (Lynch, 2003).

Successful essay examinations result from the use of authentic and explicit tasks, the understanding of the test takers and the training of the scorers as well as the ease of the scoring procedure. The essay items have to be scored in a consistent manner by implementing a deep analysis of the students' capacities from the part of the assessors. This can be achieved through the accurate application of holistic scoring.

5.3- Maintaining and Improving Holistic Scoring

The development of technology had a great impact on the field of education. The teaching and testing process which previously relied solely on human capacities has witnessed the introduction of new techniques that involve the employment of machines and computers. In 1972, a technique referred to as the scoring of the Scantron was introduced by the Scantron company which produced the first desktop scanner to read and score test answer sheets through a microcomputer.

This procedure relies on a test scoring machine that involves the utilization of machine-readable papers on which the learners answer academic test questions. This

machine scores objective tests at a rate of thirty (30) to forty (40) minutes. Also, the patented essay feature can be added to combine an objective test score with a score assigned by the teacher to the learners' essays or projects in order to get a single score (Haney et al. 1993).

Scantron highlighted the importance of computer technology in testing. In addition to this, the use of computers in schools and universities caused several changes in the teaching and testing methodologies. Computer-based technologies started to be used by the 1980's. For instance, in 1983, there was one computer for every 125 students in the United States of America while, in 2006, there was one computer for every four students.

In fact, computer-assisted teaching affected the educational context in various countries throughout the world and led to a certain improvement in the field of assessment. By 2007, many educational institutions in USA relied on computer-based testing programs. The development of psychometric techniques produced two approaches to computer-based testing. These approaches are: computer adaptive testing and automated essay scoring.

Computer adaptive testing depends on tailoring test items according to answers on previous ones. Then, the tailored items are presented to a student to locate his level of achievement. At the beginning of the adaptive test, the learner has to answer difficult items to provide a general idea of his achievement. If he answers correctly, another item that requires a higher level of achievement is presented to him, but if his response is incorrect, he has to answer an item requiring a lower level of achievement. The next item will entail a higher or lower level according to the student's performance on the initial one. To reach a stable estimate of the level of the student's achievement, a repetition of this iterative procedure is undertaken for each student.

However, this technique was criticized because the learners do not answer the same number of items. Also, students with a high level of achievement do not respond to easy items while the learners who have a low level of achievement are not required to answer very difficult items (Madaus et al., 2009).

Automated essay scoring (AES) is defined as *"...the computer technology that evaluates and scores written prose"* (Philips, 2007: 13). It is used to assess the learners written responses through the employment of the techniques of text analysis. It is a system utilizing software programs for the evaluation of writing (Madaus et al, 2009). Computerized essay scoring gives the possibility to overcome certain testing issues such as time, cost and reliability. Such a technique is mainly linked to the assessment of the learners' compositions. It aims at providing consistency to essay scoring as the variation in the scores awarded by human raters is not perceived in the scores given by e-raters (Phillips, 2007).

Generally speaking, computer-based testing is considered as a means of improving the efficiency of assessment. It can lead to tests' reliability and increase the speed with which the results are reported. This procedure is also helpful in the integration of student assessment with instruction since it can provide electronic feedback which enables the students to understand their deficiencies in writing. Thus, it gives diagnostic information about why some learners have difficulties in learning (Madaus et al., 2009).

On the other hand, some researchers argue that computer-based testing has many disadvantages which make it unable to lead to the advance of testing which considered as a technology on its own as it pessesses its own techniques and a community of practitioners called psychometricians. They prefer to use computers only for instruction. They believe that automated essay scoring may be employed as a tool to help assessors but it cannot replace expert human raters. Computer-based testing often measures attributes that lead only to better scores but they cannot lead to better writing (Phillips, 2007).

As computer based testing has certain negative aspects, the traditional ways of testing are not marginalized. Thus, human assessors are often preferable to computers. In fact, substitution of a method of assessment by another one may not have efficient results. *"Change per se is not necessarily good. If you have a stable, fair system that assesses a wide range of academic transferable skills then do not change it – although you might wish to fine tune it"* (Brown et al.,   1997: 222).

Even if the actual methods of assessment seem to have certain disadvantages, it is better to think about developing these procedures rather than changing them.

Regarding the assessment of the writing skill, one can state that the direct methods are often seen as the most efficient instruments used for such a task.

> *Within the community of writing teachers and instructors, the task of evaluating writing and providing quality feedback is viewed as a complex process, and many within that community believe writing cannot be evaluated appropriately by computer assessment programs.* (Philips, 2007: 15)

This means that the assessment of writing depends on human raters whose judgement is essential in determining the degree of the learners' performance. Among the methods used for rating essays, holistic scoring is regarded as the most suitable and efficient one since it provides a general judgement about the quality of the students' compositions.

Holistic scoring considers writing as a unitary trait that cannot be rated in a mechanical manner (Lynch, 2003). This approach regards the writing skill as a set of interrelated features that have to be evaluated in an integrative way. *"The overall quality of the response is valued more than the discrete elements of the response; the primary advantage of holistic scoring is the simultaneous considerations of all components"* (Welch, 2006: 313).

Thus, the performance of the learners is assessed as a whole by judging how well the students manage to combine their content knowledge with the different language elements to produce a meaningful piece of writing.

This type of scoring can lead to the acquisition of scores that are fairly consistent whether the piece of writing is read several times by the same assessor or it is read by various raters. At certain times, these scores may be approximately identical. The teachers' personal judgement can be very efficient especially when they are accustomed to the task of assessing essays holistically.

> *Holistic scoring tends to be quite reliable. With a little training, different people are likely to rate assessment items (papers, projects, etc.) the same. In addition, the same person is likely to rate the same item similarly each time she or he examines it.* (Colburn, 2003: 40)

Therefore, this procedure can provide a better way of assessing the learners' writing performance by focusing on the general quality of the students' production. The reliability of the rating procedure may be affected by the scoring of novice raters who tend to put the emphasis on grammatical accuracy (Weigle, 2002). However, these teachers can develop their skill in scoring consistently by acquiring more experience as well as by reflecting on holistic rating and discussing with other teachers to reach an agreement concerning the general guidelines that correspond to this technique. Also, reliability often increases when each paper is read by two teachers or more (Hyland, 2003).

Another advantage of holistic scoring is its capacity to reduce the time of evaluation as it offers a certain speed in assessing compositions. The rate of scoring of trained assessors is one page a minute (Finn, 1997).

Generally speaking, rating is subjective. Thus, variations in the scores given by different scorers will always exist. Some researchers *"...attempt to see disagreement between raters as the source of valid information about candidate performance rather than as an indication of unreliability and a source of unfairness"* (Lynch, 2003: 59). This can lead to the conclusion that this kind of rating possesses validity as it judges the learners' writing performance as a complex skill.

Holistic scoring can be developed by establishing clear assessment criteria that can help the teachers in producing a general judgement of the learners' writing performance. Jones et al.(1999) suggest four main categories of assessment criteria which are: overall impression, content, organization and language. The overall

impression means that the essay reflects the students' understanding of the given task. The content refers to the learners' ability to employ the information input and knowledge to elaborate the ideas and provide details that elicit the issue discussed by the topic. Organization is linked to the textual layout of the essay. Language refers to the accuracy in the employment of lexical and grammatical patterns as well as the fluency and appropriateness of the written discourse.

The use of such criteria will help the assessors to understand how the learners combine various features to produce a text. Fairness in scoring can be obtained through the raters' general agreement upon the evidence of the existence of these criteria within the students' written product as they constitute the critical features that shape the scorers' judgement of the learners' performance (Lynch, 2003). The function of assessment criteria is to provide a reliable and valid indicator of the quality of the students' essays (Turner and Street, 1999).

Therefore, the employment of holistic criteria which include the features of overall impression, content, organization and language helps the raters to score the learners' essays by taking into consideration the students' ability to achieve successful written communication via the use of their knowledge of the subject and their mastery of the target language.

Holistic scoring can become more efficient by training teachers to rate essays consistently. Since this technique relies on providing an overall response to the compositions, the assessors have to be trained to respond in the same way to the students' written product (Hyland, 2003).

Rater training sometimes takes place through a procedure called the moderation meeting which involves various steps. First, the scorers are required to assess samples of performances independently. Next, the scores are compared with the ratings that have been formulated in advance by experienced assessors. There is a discussion concerning the differences between the scoring of the raters-in-training and the experienced raters. Then, instructions are given to the novice scorers in order to moderate their way of scoring and bring it into line with the one of the

experienced raters. Finally, the raters-in-training are required to rate real essays of the students. During this phase, they are monitored by the experienced assessors to make their rating more consistent (Lynch, 2003).

This kind of counselling can help the raters to produce a qualitative judgement of the learners' compositions on the basis of a shared understanding and interpretation of holistic scoring. Training teachers on holistic scoring enables them to become skilled in reading the learners' composition with a preconceived view about the learners' interaction with the topic and the writing processes as well as the expected quality of the written product. In addition to this, co-ordination meetings can lead writing examiners to agree on a standardization of the main writing features that should govern the students' essays (Shaw and Weir, 2007). In this way, they will have the opportunity to make decisions concerning writing standards that provide a basis for the holistic judgement of the learners' compositions.

Another strategy of scoring essays relies on the utilization of analytic marking scales.

5.4- Employing Analytic Marking Scales for Scoring Compositions

Scoring essays is subjective since it involves the teacher's appreciation and point of view concerning the learner's answer. In order to decrease the degree of subjectivity, analytic scoring may be employed. This kind of marking involves the use of various components in order to assess the students' essays.

> *Holistic scores represent teachers' overall impressions and judgement... In analytic scoring, different components or features of the students' responses are given separate scores (on an essay, spelling, grammar, organization and punctuation might be scored separately).*

> (Genesee and Upshur, 1996:206)

Analytic scoring scales may be used to rate essays and compositions for the sake of providing feedback to the learners about their performance. In this way they will get information about their level of achievement and their weaknesses.

Analytic scales include component scores which not only represent numerical values but have also a meaning that is directly linked to identifiable text features and explicit performance standards. Thus, they help the teachers to score essays by focusing on the existence of the established criteria in the evaluated composition (Ferris and Hedgcock, 2004).

Analytic scoring scales are employed to gain a series of separate scores for separate subskills by rating the learners' written product on various dimensions like organization, ideas, spelling and mechanics. They are in the form of general guidelines that represent a model of a skilfully organized and segmented piece of writing (Babin and Harrison, 1999).

Luongo-Orlando (2003) suggests an analytic rating scale for scoring essays that includes six writing criteria. These criteria involve the writing process, content, organization and development, visual presentation and format, language usage and the application of language conventions. For each criterion there are five levels. The first level whose descriptor is 'excellent' is awarded five points. The second level is assigned four points and its descriptor is 'good'. The third level takes three points and it is described as 'satisfactory'. The fourth one refers to a 'limited' level that is given two points. The descriptor of the last level mentions that it 'needs improvement' and its score is one point.

The following writing evaluation sheet (adapted from Luongo-Orlando, 2003) can be used to score essays in an analytic way.

Criteria	Rating scale				
	5 points	4 points	3 points	2 points	1 points
Content: - Clear ideas - Accurate information - Providing explanations - Adequate details					
Organization: - Logical development - Structured essay					
Word choice: - Rich vocabulary - Effective word choice					
Language conventions: - Knowledge of spelling - Mastery of grammar - Use of capitalization and punctuation					

Table 5.2 - Writing Evaluation Sheet

The above evaluation sheet is based on the assessment of four writing criteria which include the content of the essay, organization, word choice and language conventions. The score of each criterion is five points.

In fact, analytic scales can supply the teacher with more details about the learners' performance. Their employment gives the possibility to test particular qualities of the written product. However, they are time-consuming and require more concentration from the teachers who have to read the essay through various dimensions so that to be able to judge the learners' good command of each criterion

stated in the scoring scale. This why scoring rubrics are considered as efficient rating instruments.

5.5- Constructing Scoring Rubrics

A rubric is a scoring scheme that employs a set of criteria for the evaluation of the students' performance. It provides a description of each level of performance that the learner is assumed to achieve.

Rubrics are classified into two categories. The first category is characterized by the type of scales. It includes holistic and analytic scoring rubrics. The holistic scoring rubric incorporates various performance criteria in a single descriptive scale that requires a general judgement of the students' work. The analytic scoring rubric gives the possibility to make a separate evaluation of each aspect of the learner's performance through multiple scoring scales involving various criteria.

The second category of rubrics is identified on the basis of the task features. It is divided into general and task specific scoring rubrics. The general scoring rubric is utilized for the evaluation of a certain task without focusing on a particular context. The task specific scoring rubric is employed to assess the learner's performance on a special assignment (Elfner, 2005). Therefore, these categories lead to the classification of rubrics into four types which are: holistic-general scoring rubrics, analytic-general scoring rubrics, holistic-task specific rubrics and analytic-task specific rubrics. The following table (adapted from Elfner, 2005) provides a description of these types.

Type of Rubrics	Description
Holistic-general scoring rubric	-Employing only one scale -Evaluating a task without giving importance to a particular context
Analytic-general scoring rubric	-Multiple scoring scales -Assessing an activity without focusing on a specific context
Holistic-task specific rubric	-Using a single scoring scale -Putting emphasis on a learner's performance in a particular context.
Analytic-task specific rubric	-Utilizing multiple scoring scales -Focusing on the learners' performance on a single assignment

Table 5.3- Types of scoring rubrics

The choice of the kind of the rubric that should be employed for scoring essays depends on the type of tasks involved and the purpose of assessment.

To be efficient tools of measurement, rubrics must possess four essential features. These characteristics are: content, clarity, practicality and technical quality.

Content which is also referred to as coverage, means that the rubric really indicates the main features of the learners'performance. Clarity, or detail, is the explicitness of the rubric. Practicality refers to the usefulness of the rubric to the process of assessment. Technical quality, which is called soundness or fairness, implies a certain agreement on the scores from the part of the raters so that to ascertain that this rubric is fair to all the learners (Arter and McTighe, 2000).

The development of scoring rubrics goes through different steps. First, the teacher has to decide about whether to use an analytic rubric or a holistic one. Analytic rubrics may be used in the case of formative assessment to give information about the process of teaching and provide feedback to the learners. On the other hand, holistic rubrics are more suitable for the assessment of the students' performance.

The next step in rubric development is the identification of the performance criteria which may include organization and structure, level of understanding, ideas, coherence of presentation, knowledge of the demonstrated material, mechanics, voice, creativity and degree of task completion (Elfner, 2005).

Most of the time, rubrics used for assessing writing focus on six criteria referred to as the six traits of writing involving ideas, organization, voice, word choice, sentence fluency and conventions. The criterion of ideas means the content of the essay. Organization refers to the logical presentation of ideas and the structure of the written product. The voice is the degree of the conviction of the writer that is reflected through the use of words. The word choice is embodied in the richness of the language. Sentence fluency means the rhythm and flow of the language. Conventions refer to the mechanical correctness of the essay which includes spelling, punctuation grammar and usage (Arter and McTighe, 2000).

After establishing the assessment criteria, levels of performance or performance standards should be designed for each criterion. These levels are expressed in the form of descriptors. They may range from two to five levels of performance for each criterion. The presentation of the description of these levels may take the following forms: inappropriate - appropriate, incomplete - complete, incorrect - partially correct - correct, vague - emergent - clear, ordinary - interesting-challenging, unsatisfactory - satisfactory - outstanding, needs improvement - adequate - proficient - exemplary (Elfner, 2005). Each level of performance is represented by a score that indicates the degree of achievement. Sometimes, a score rubric may be converted into a letter grade. The following rubric (adapted from Luongo – Orlondo, 2003) may be used as a tool for scoring essays.

Criteria	Levels of Achievement			
	Level 1	Level 2	Level 3	Level 4
Content	- Few details related to the topic	- Some details related to the topic	- Most of the details are related to the topic	- All the details are related to the topic
Organization	-Limited content development -Few ideas flow logically	-Minimal content development -Some ideas flow logically	-The content is fairly well developed -Most of the ideas flow logically	-Fully developed content -Logical flow of all the ideas
Sentence fluency	Limited language rhythm	Minimal language rhythm	The sound of word patterns is fairly well elaborated	The sound of word patterns is consistent
Vocabulary	- Limited vocabulary -Lack of clarity and precision of word choices	-Appropriate vocabulary -Sometimes including expressive word choices	-Clear and precise vocabulary -including expressive word choices most of the time	-Clear and precise vocabulary -Consistent word choices
Spelling, grammar, punctuation	-Containing several errors	Containing some errors	- Containing few errors	Containing very few errors

Table 5.4 - Writing rubric

The above rubric comprises five criteria. Each criterion includes four levels that are assigned a score of four (04) points. Thus, each level is awarded one point.

In fact, the use of rubrics facilitates the scorer's task and gives detailed information about the learners' performance. Scoring rubrics can provide feedback to the students about their strengths and weaknesses. They may be employed as instructional instruments that help the learners to understand the meaning of writing performance criteria in order to give them the possibility to improve their levels of achievement (Arter and McTighe, 2000). In this way, the students may practise writing essays that include most of the criteria that are involved in the process of assessment.

Therefore, rubrics can have two essential functions. First, they represent a tool of measurement that is utilized to make a judgement about the learners' performance. In addition to this, they may blend testing and instruction through the use of the assessment results to develop the students' writing skill. In fact, criteria can also be useful in the assessment of dissertations.

5.6- Strategies for Assessing Dissertations

Dissertations reflect the students' degree of achievement and ability to formulate information in a written product. They provide evidence about the learners' mastery of the target language and their knowledge of the content of the topic as well as their way of applying research methodology. Thus, specific criteria should be examined when scoring projects. Mauch and park (2003) propose the use of an evaluation form for the assessment of dissertations relying on forty characteristics. These criteria revolve around eight general divisions including the topic, language mastery, the research questions and hypotheses, the review of literature, data collection, data analysis, data interpretation, and the conclusions. The following dissertation evaluation form (adapted from Mauch and Park, 2003) can be employed to assess projects.

- Name of the evaluator: - Title of the dissertations: - Name of the student:					
Characteristics being evaluated	Evaluator's remarks				Evaluator's notes on the rated item
	Poor	Mediocre	Good	Excellent	
1- The title is clear and concise					
2- The problem is clearly stated					
3- The hypotheses and research questions are clearly stated					
4- The literature review is efficiently summarized					
5- Important terms are well defined					
6- The sample population is described					
7- The data gathering methods are described					
8- Appropriate methods are used to analyse data					
9- Tables and figures are used effectively					
10- Results of the analysis are presented					
11- Suggestions for improvement are logical					
12- The conclusions are clearly stated					
13- A suggestion for further research is cited					
14- The bibliography is well organized					
15- Correct sentence structure					
16- Correct spelling					
17- Correct punctuation					
18-Minimum of grammatical mistakes					
19-Minimum of typographical errors					
20-The material is clearly written					
General Comments	The mark: /20				

Figure 5.1- A Dissertation Evaluation Form

The advantage of the assessment of dissertations is the double or multiple marking as a process of triangulation often exists through the participation of internal and external examiners in addition to the supervisor.

However, the assessment of such projects may involve subjectivity at least from the part of the assessor who has been supervising or advising the candidate. Also, it can be over-dominated by matters of style and structure (Race, 1998). The scoring of dissertations is a difficult task.

> *Three common difficulties in the assessment of projects are variations between assessors, variations in the difficulty of projects and the variations in students' and supervisors' contribution to the project report. One way to reduce variability of assessors is to use an agreed set of criteria.*
>
> (Brown et al. 1997 :127)

In order to decrease the assessors' subjectivity and variations in rating, specific criteria should be established to judge the students' written work. These criteria may involve the organization of the project, the methodology employed, the learner's knowledge of the topic discussed and his/her mastery of the language. In addition to this, raters' training can produce reliability. Also, the use of assessment results as feedback may promote the testing situation.

5.7- Assessment as Feedback on Performance

Feedback has a great effect on students' learning and achievement. It allows the process of assessment to become a central part of the learning/teaching process. Feedback is defined as a source of *"...giving information about the gap between the actual performance level and the reference level"* (Gomez and Osborne, 2008:41). In fact, feedback is employed in order to help the learners to reach an acceptable level of performance by being conscious about their weaknesses. It leads to clog the gap between the actual or current performance level and the reference or standard level. The assessor's role is to raise the students' awareness about the relationship between learning and assessment through feedback on the learners' work represented by the awarded marks as well as the teachers' comments and correction of the learners' mistakes.

The marking process can provide students with feedback on their performance. When assessing essays and compositions, the teacher not only awards a mark for the piece of work but he also indicates that there are errors and corrects them whether directly or indirectly. However, the use of direct correction which involves writing the correct version on the students' papers may lead the teacher to waste time. Thus, it is preferable to employ the indirect correction which implies circling the errors, underlining them or using symbols without writing the correct words. Nott (2008) suggests a set of symbols that refer to the category of the students' errors. They are summarized in the following table.

Symbols in the margin or between the lines		Symbols in the text	
symbols	Their meaning	symbols	Their meaning
SG	Should be singular	_____	Unacceptable
PL	Should be plural	/	Repetition of similar error made previously
AG	Make verb ending agree with subject	(_____)	Omit underlined words
SP	Error of spelling	- - - - - - - - - -	Not exact meaning
V (or Voc)	Error of vocabulary	~ ~ ~ ~ ~	Awkward style: the phrase needs to be reformulated
T	Error of tense	. . . .	Acceptable, but not the most appropriate word
G	Error of grammar	/ /	(slashes before and after suggested wording): more appropriate wording
WO	Error in word order		Good choice of word; good use of language
ME	Meaning not clear		
( )	Omit bracketed words.	✓	
^	Words missing		

Table 5.5 - Symbols used for indirect correction

These symbols should be explained to the learners so that they can identify their mistakes once their assessed pieces of work are handed to them.

In this way, assessment gives the learners an idea about their deficiencies so that to motivate them to learn. The information supplied about how the students performed in examinations represent feedback that affects the students' learning and leads them to attempt to adjust their performance to the standard level in future

assessment situations. Feedback is sometimes referred to as reinforcement. It can serve various purposes as it can be corrective, explanatory or diagnostic. It may be verbal or written since the teacher can give oral remarks to the learners or write comments on their work.

Various kinds of feedback exist. However, in the assessment of writing the most common types are evaluative and descriptive feedback. *"Evaluative feedback indicates how good the action taken is, but not whether it is the best or the worst action possible"* (Sutton and Barto, 1998:25). Thus, it gives an idea about the level of performance formulated in numerical scoring or letter grades. Evaluative feedback is also called judgemental feedback as it provides the students with a judgement of their writing through the marks they receive.

Descriptive feedback which is referred to as Criterion-based feedback or reader-based feedback focuses on the learners' strengths or weaknesses since it describes the quality of writing by showing the positive points in the written work and the areas needing improvement (Stiggins et. al, 2004). This can be achieved through the assessor's written comments on the learners' prompts.

> *Teachers can provide descriptive, "reader-based" feedback in which they describe how they as readers are responding to or processing students' writing. In doing so, they provide writers with a description of their processes in reading a draft.*
>
> (McArthur et.al, 2006: 226)

This means that this type of feedback informs the learners about their mistakes and trains them to correct them so that they can achieve self assessment which leads them to be conscious about their degree of performance and their deficiencies.

Descriptive feedback is given more importance than other types of feedback and is often used by assessors as it makes *"…students know where they are going, where they are now, and how to close the gap between the two"* (Stiggins et. al, 2004:40).

Therefore, the testing process is not limited to the act of awarding marks as it can be used as a means of developing the learners' performance by eliciting their weaknesses through the use of descriptive feedback. The provision of feedback can also take place through portfolio assessment.

5.8- Portfolio Assessment

A portfolio is a collection of the learners' written work extended over a period of time. It is often used as an instructional instrument but it can be employed for assessment too. Portfolio assessment attempts to demonstrate students' progress in writing. The success of such a procedure relies on the collaboration of the teacher and students who have to reach the objectives set up for the writing tasks. *"Portfolio assessment takes into account the need for the generation of ideas and thoughtful revision over time"* (Lippman, 2003: 201).

Therefore, sufficient time should be devoted to the explanation of the procedures involved in assembling portfolios to the learners as well as the provision of appropriate feedback on the students' written work at different stages of the written process.

Portfolios must be linked to the objectives of testing. They may progress through a semester or the whole year according to the assessment purposes they fulfil. They may spread over several years of study (Barone and Taylor, 2005). Clear procedures have to be set up for the grading and evaluation of portfolios. This is why the teachers should agree on the structure of the portfolio and assessment criteria to be able to assess the students' work in a fair and meaningful way (Weigle, 2002).

In order make portfolios successful, the teacher has to discuss with the learners about their weaknesses and provide suggestions that can help them to improve their grades after rewriting their work.

In fact, portfolio assessment may serve to evaluate program effectiveness or student progress. Such a process can achieve these purposes if it is based on specific strategies such as deciding about the tasks involved in the portfolio, establishing a set of criteria for evaluation, ensuring scoring consistency and using the results of portfolio assessment effectively (Urquhart and Mciver, 2005).

Portfolios help the teachers to get a broader idea about the learners' capacities to give them the possibility to develop them. Also, it enables the students to monitor their achievement and learn more about the writing skill through rewriting, revising and proofreading their written products.

Portfolios may be helpful for the assessment of the learners since they represent a collection of written works that display a range of writing performances. They allow the students to reflect on their production and select the pieces of writing that should be included. The main characteristic of portfolios is delayed evaluation which allows the learners to revise their work before presenting it. Therefore, portfolio assessment provides the opportunity to get more information about the students' writing abilities since it is concerned with a collection of writing samples.

The organization of a portfolio has to rely on three features which include collection, reflection and selection. Thus, pieces of writing are collected. Then, the learners reflect on the content of the essays, their purpose and how they reveal the students' progress. Finally, they select the pieces of writing that should be employed for assembling a portfolio (Weigle, 2002).

In addition to portfolios, distributive evaluation may be employed to develop the process of assessment

5.9- Distributive Evaluation

Distributive evaluation is a method that can be used for assessing the learners' essays by various teachers or readers. It is often utilized in the assessment of students' electronic portfolios which may be read by several advisors and educators. *"Distributive evaluation involves seeking multiple perspectives and responses to student compositions"* (Whithaus, 2005: 88).

This procedure implies the incorporation of distinct forms of appreciations of the written product from the part of multiple audiences including teachers and administrators. Each evaluator is requested to apply his own expertise and judgement to come out with an accurate evaluation of the learners' essays.

Therefore, this technique does not compel the readers to produce similar opinions about the students' prompts. It focuses on the provision of multiple and distinct responses to the learners' compositions to ensure the validity of the assessment and to increase objectivity. There is no specific rubric that is imposed on these readers who have to rely on their personal judgement to evaluate the written work (Borton et.al., 2009).

Another purpose of this system is to agree on a set of criteria or standards of performance for the evaluation of the students' essays. In addition to this, distributive evaluation allows the learners to get sufficient feedback on their work through the different evaluations provided by the audiences.

This procedure may also be used in the classroom to teach performance criteria to the learners. In such a situation, the teacher asks the students to read the work of their peers. In fact, the learners do not grade each other's composition. They just make a discussion with the teacher concerning the different pieces of writing and try to comment on the way the given criteria may be respected when producing essays.

Therefore, distributive evaluation can help the assessors in improving their views about the concept of writing assessment. It can produce collaborative learning too.

> *For the teaching of composition, this multiple interactivity urges moving beyond a student – teacher vision of the composing process and to see the student within a framework of other students, the teacher, and potential audience members outside of school.*   (Whithaus, 2005: 49)

This means that distributive evaluation enables the teacher and the students to get insightful remarks and ideas about the process of writing in order to make more efforts to improve the teaching and testing situation. This technique can be useful to the learners who will have a variety of responses that reflect their strengths and weaknesses. Thus, the students will have the opportunity to play an active role in the process of assessment (Borton et al, 2009).

Distributive evaluation may lead to a better perception of the process of writing and evaluating compositions in order to create more developed methods for the teaching and assessment of writing. Another strategy that is also beneficial is known as product analysis.

5.10- Product Analysis

Product analysis refers to the assessment of students' performance by examining students' written products like assignments and portfolios. This procedure can involve the evaluation of the quality of the written work produced by each learner or a comparison of students' products. *"Product analysis is not a single approach but a set of informal techniques"* (Cline, 1992:163). These techniques or strategies include examining completed assignments, observing the patterns of the written work and evaluating the learners' degree of success in using learning strategies. *"Analysing the quality of a product can be facilitated by the use of specified criteria. These criteria are used to create a product rating scale"* (Spector, 2007:781). The main criteria employed for analysing students' products are the organization, originality and persuasiveness of the written work in addition to the

task comprehension. These features are considered as the main elements that constitute the various kinds of product rating scales that are involved in the evaluation of students' writing.

Product rating scales can take one of three forms which include numerical scales, descriptive scales and checklists. Numerical scales incorporate a range of sequential numbers defined by particular labels. Thus, each criterion is awarded a numerical value in terms of a specific order.

Descriptive scales resemble the numerical ones. However, they put more emphasis on verbal statements. Each statement can be assigned a given number.

A checklist provides a list of the components that must exist in a the learner's product or performance.

> *Checklists and rating scales can be used in evaluating the*
> *work process and products of individuals... these techniques*
> *help diagnose strengths and weaknesses in the ways in which*
> *a student goes about his work and in the quality of the*
> *completed work.*   (Arora and Vashist, 2004: 216)

Therefore, the analysis of students' written product can provide information about their performance in a continuous way so that to enable the learners to achieve better results in the examinations.

The main purpose of such a procedure is to identify the strategies used by the learners to produce a piece of writing and to assess their strengths and weaknesses in order to decide about the instructional changes needed for the improvement of the learners' writing abilities.

5.11- Using the Six Traits for Teaching and Assessing Writing

The assessment of writing is closely linked to the way this skill is taught. In fact, the improvement of the learners' writing performance is interrelated with their degree of understanding of the nature of this skill and the qualities of a good composition. When teaching writing, the teacher has to give the learners the opportunity to develop their writing capacities and to produce essays that are acceptable at the level of form and content.

Many researchers agree on the necessity to use the six writing traits to enable the learners to generate compositions. In 1961, Paul Diedrich and his colleagues John French and Sydell Carlto conducted a study that led to the conclusion that the writing features into five categories of these criteria included ideas (relevance, clarity, development), form (organization), flavour (style, interest, sincerity), mechanics (grammar, punctuation, spelling), wording and phrasing (choice and arrangement of words). In the 1980's, Diedrich's list of criteria was extended. These writing features that are known as the six traits are widely used. They are classified as follows: ideas and content, organization, voice, word choice, sentence fluency and conventions (Rog and Kropp, 2004). In the 1990's, these elements were expanded by adding another criterion referred to as presentation. These scoring criteria are called the 6+1 traits (Culham, 2003).

The teachers can teach their students the six traits in order to develop their writing competence. First, each element can be introduced on its own. The trait of ideas and content can be developed by making the learners practise how to elaborate the main idea of the piece of writing and supply strong details to enhance it. Since the feature of organization which refers to the logical presentation of ideas is linked to the purpose of writing, the students can learn about how to organize their essays by teaching them about the various genres of texts. The criterion of voice is difficult to define and teach as it is connected to the writer's feeling and style. Thus, the teachers have to instruct their students about how to provide voice in their writing by focusing on reflecting a specific tone for each distinct audience.

The trait of word choice can be taught by urging the learners to read a variety of texts that will enrich their vocabulary stock. Sentence fluency which refers to how a text sounds can be achieved by making the learners produce an essay that includes sentences that vary in their length, type and structure.

Conventions, or the mechanics of writing, are necessary for a good composition. They can be taught through lessons that focus on the grammatical elements that constitute certain difficulties for the learners. Also, the use of various drills about capitalization, punctuation, spelling and grammar can help the students to master the target language and become able to respect the trait of conventions in their pieces of writing (Rog and Kropp, 2004).

Trait-based writing enables the learners to improve their abilities as they have the opportunity to understand the qualities of a good composition and practise writing essays according to the principal features that should exist in a written product. It also helps the students to become self-evaluators (Culham, 2003).

Teaching the six traits to the learners gives them the possibility to achieve higher levels of proficiency by allowing them to make a revision of their writing for the sake of identifying their strengths and weaknesses.

Rog and Kropp (2004) propose six rubrics that can be used to teach writing traits. Each rubric includes five descriptors that are linked to a specific trait. For instance, the descriptors of rubric of ideas and content describe the students' levels of performance concerning the development of the theme, the clarity of ideas and the appropriateness of the supporting details.

The descriptors of organization focus on the structure, the unity, the introduction and the conclusion of the piece of writing. The rubric designed for voice puts the emphasis on the student's energy, personality and sense of the audience.

The standards of the trait of word choice are described in terms of the degree of the precision, clarity and effectiveness of the employed vocabulary.

The levels of the learners' performance as far as sentence fluency is concerned are based on rhythm, tone and variety of the utterances that form the essay.

The rubric of conventions includes descriptions of the correctness or incorrectness of the use of capitalization, punctuation, spelling and grammar. These rubrics can be set by the teacher or they can be constructed by the students themselves in order to achieve a process of self-evaluation.

Since teaching and testing are interrelated, the instruction of the six traits through separate rubrics aims at preparing the learners for the examinations which will be evaluated via such features. The employment of these six traits for the assessment of writing helps the teachers in making a judgment of the students' essays through an understanding of the characteristics of good writing. This procedure ensures the accuracy and consistency of scoring. It also gives meaningful feedback to the learners through the comments that are supplied for the performance of the students in each trait (Culham, 2003).

The evaluation of the learners' composition can rely on the use of six rubrics or one rubric including all the six traits. Each trait is assigned a score from one to five points. The total mark is obtained out of thirty (30) points. Rog and Kropp (2004) propose a scoring rubric that is based on the six traits of writing. The following table (adapted from Rog and Kropp, 2004: 45) summarizes the descriptions they provided for the different descriptors related to each trait.

	5 Exemplary	4 Strong	3 Satisfactory	2 Developing	1 Below expectations
Ideas/ content	Well - developed theme	Clear and focused	Evident main idea with some support	Some attempt at support or expansion	Writing may lack a central idea
Organization	Effectively organized in logical and creative manner	The structure moves the reader smoothly through the text	Organization is appropriate but conventional	Introduction and/or conclusion are not well developed	Ineffective or non-existent lead and closure
Voice	Individual and engaging	Sincere and engaging	Pleasant tone; appropriate voice	Voice may be erratic or non-existent	Flat, lifeless
Word choice	Engaging words, carefully selected to convey the intended impression	Broad range of vocabulary	Words are adequate and correct	No evidence of precision or description	Limited vocabulary range
Sentence fluency	Well-crafted sentences with varied length and structure	Good variety in sentence length and construction	Occasional effort at variety and fluency	Lacks variety in sentence length and construction	Choppy sentences
Conventions	Mastery of writing conventions	Strong control of conventions	Reasonable control over conventions	Frequent errors in conventions	Numerous errors in conventions

Table 5.6 - A scoring rubric for writing

The utilization of the six traits for teaching and assessing writing can have a positive effect on the learners' perception of the characteristics of acceptable compositions which will lead them to improve their command of each trait so that to produce better essays. This technique may also help in teaching and assessing the learners' performance in specific genres of writing. In addition to this, the use of these traits enables teachers to link the instruction of writing to the objectives of assessment. Thus, the improvement of the teaching of this skill will result in the development of the learners' level of writing. The teaching of writing can be enhanced via assessment too.

5.12- Assessment as a Means of Teaching Writing
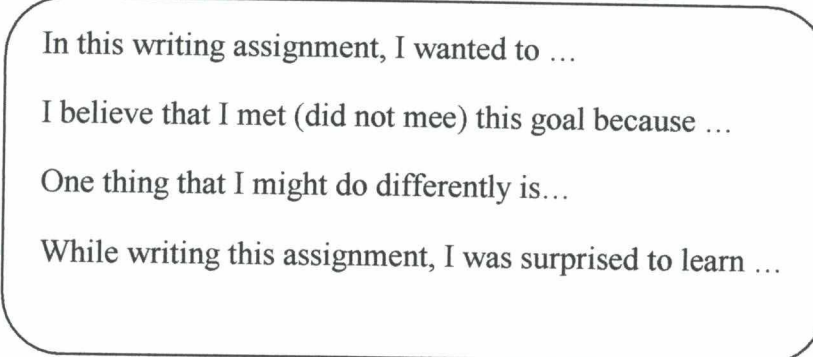
Teaching writing requires great efforts from the part of the teacher since it takes time to make the learners good writers. The students can learn more about when being assessed as they have the possibility to display the amount of knowledge they acquired about the target language. Urquhart and Mciver (2005:27) refer to *"...assessment as writing - to -learn"*. Certain learners do not practice writing during the teaching sessions. Thus, the essay examination is an opportunity for them to produce pieces of writing.

The learners will have the possibility to practise writing during the examination. After having seen the results of the examination, they may produce a different draft concerning the same task and compare between their performance in the two situations. *"Feedback on the content of learners' writing can do a lot to increase the amount of writing that learners do and to improve their attitude to writing"* (Nation, 2008: 137).

The examination results will give the students an idea about their degree of performance and lead them to think about improving their level. In addition to this,

the teacher can use assessment as a tool to diagnose learners' difficulties in writing so that to make remedies and adjust the teaching of writing to the students needs.

Urquhart and Mciver (2005) state that effective assessment should be based on four traits including clarity, variety, sound pedagogy and reliable research. Clarity means understanding what is being assessed. Therefore, both the teacher and students have to be aware about the purposes of testing writing. Variety refers to the use of multiple samples for assessment that involve different types of writing including narrative, expository, argumentative and descriptive writing. Sound pedagogy means the use of strategies that lead assessment to reinforce the teaching practices. For instance, the teacher assesses the students' writing and asks them to reflect on their own writing by asking them to complete a goal sheet as the one that is provided in the following figure (adapted from Urquhart and Mciver, 2005).

In this writing assignment, I wanted to …

I believe that I met (did not mee) this goal because …

One thing that I might do differently is…

While writing this assignment, I was surprised to learn …

Figure 5.2 – Student's goal sheet

The use of such a tool will help the students to think about their writing by encouraging them to monitor their performances. Finally, reliable research enables teachers to perceive the importance of assessment by urging the learners to write and reflect on their own writing.

In fact, the instruction of writing can be develop through the assessment results which may be an important factor in motivating the learners.

5.13- Raising Students' Motivation through Assessment

Motivation is an essential element in developing the learners' performance. According to Stiggins et.al, *"...accuracy + descriptive feedback + student involvement = achievement gains"* (2004: 38). Thus, achievement gains in students' performance can occur via the accuracy of the results of assessment in addition to descriptive feedback and the students' collaboration in the process of learning.

This means that there is a close relationship between the learners' motivation and their degree of achievement. This motivation can be enhanced through testing which provides information about the students' performance. *"Teachers can use assessment to foster motivation to learn, not just to pass the test"* (Maxine Earl, 2003: 76). The results of examinations can be used to push the students to make more efforts by explaining to them that the marks they obtained are not a sign of reward or punishment but they are reflections of their performance which may change according to their degree of willingness to improve their level.

The students' motivation can be raised by providing them with frequent feedback on performance, enabling them to use self-assessment accurately and giving them learning tasks related to everyday life (Stiggins et. Al, 2004). Therefore, the learners become motivated when they feel responsible for their work and have a sense of control over their learning. Also, descriptive feedback leads them to think about developing their writing skill and avoiding the mistakes they made in previous examinations.

> *When assessment is integrated into the learning process, students and teachers can work together with a shared expectation of finding out what makes sense. This can unleash students' natural curiosity and encourage them to engage in order to acquire knowledge or skills.* (Maxine Earl, 2003: 76)

Testing is related to learning as the former cannot take place without the existence of the latter. In the same way, learning is related to the students' motivation. This means that motivation has an effect on the testing situation. Thus, the learners who do not have an incentive to perform well in examinations will not try to progress. The students have to be aware about the importance of trial and error in achieving success. Their results in the examinations must become a source of information about their writing performance. A collaboration between the learners and the teacher is needed too.

In order to make assessment an instrument of motivation, certain strategies must be employed such as making assessment more accurate by including tasks that capture students' imagination and knowledge. Assessment should provide information to the learners about their writing abilities so that to become a motivator for them. Also, it should scaffold learning (Maxine Earl, 2003).

In fact, essay examinations involve tasks that can give the students a detailed picture about their way of presenting and organizing knowledge through the use of the target language as well as informing them about their weaknesses.

In this way, the learners' motivation is enhanced via holistic assessment whose results may be utilized by the teacher to encourage the learners to improve their writing skill. Also, the strategies employed by the teachers to teach writing may have a great impact on the students' writing performance.

5.14- The Use of a Variety of Techniques to Teach Writing

Essay writing is a very important activity that enables the learners to express themselves and leads them to learn the language in a better way. According to Tynjälä (2001), there are two main approaches to learning which include the reproductive and the constructivist view. The reproductive outlook requires factual recall of information while the constructivist view is a current approach that focuses on making the learners use the language to produce texts that formulate their understanding of the taught data. Thus, the constructivist activities rely on reading,

writing and learning. Essay writing is seen as the most suitable activity that can help the students in their learning process as it deals with discourse synthesis which involves selecting, connecting and organizing information.

Although teaching compositions is a useful tool for learning, it remains a difficult task for the instructors who have to take decisions concerning the techniques they can employ to teach writing in an effective manner.

Many strategies can be employed in order to enhance the instruction of this skill. One of these strategies is to use reading to teach writing.

Some researchers think that these two skills are inseparable since reading and writing abilities are interrelated.   In the 1980's, researchers claimed that these two skills helped the students to learn more effectively. They also started to think about how to exploit reading and writing relationships to develop the teaching of composition. Reading constitutes models of good writing and can provide the learners with ideas that develop the three necessary foundations of writing which include thinking, discussion and responses (Vandrick, 2003). The teachers' understanding of such connections can guide them in designing courses that promote writing and literacy.

The students may be asked to utilize texts resources for writing activities like summarizing or discussing the content of the text. However, this can be a difficult task for the learners as they have to decide about the amount of information that should be taken and the mechanisms that should be employed for the transformation of the textual information into an essay. The learners have to make more effort as such activities urge them to use various strategies that rely on critical decision making, planning, processing and revising. However, continuous practice can lead to satisfactory results (Grabe, 2003).

The main principle that links reading to writing is the idea that these skills are a form of interaction between the author and the audience. The practise of reading and writing together leads the learners to assume the roles of authors and audiences so that they can produce good essays. Thus, "..._reading and writing as interaction' highlights the notion of dialogue at a distance through the medium of the written text. Research studies have shown that envisioning an audience can improve writing_" (Grabe, 2003: 245). Therefore, reading different materials and using the information of these texts to produce essays enable the learners to develop their writing levels as these two skills share some common processing abilities.

Reading takes the form of a serial process that involves graphic input and visual encoding as well as higher order processes such as the intervention of syntactic and semantic interpretations (Sadoski and Paivio, 2001). On the other hand, writing involves thinking, organizing and linking, in addition to the manipulation of various levels of language. The knowledge of morphology, syntax and semantics is involved in reading and writing. This means that the reader memorizes vocabulary and syntactic structures that can be employed in the writing process.

In addition to the use of the technique of reading a variety of texts, the employment of literature for teaching writing can be very beneficial to the learners. Since literature is an instrument of entertainment and education, it can provide the learners with the possibility to expand their vision of the world by reading stories that introduce new experiences and cover universal themes such as morality, work and family.

> _When reading is enjoyable, pleasurable, it arouses interest_
> _and a sense of connection and in turn motivates students to_
> _respond in discussions and in writing. This kind of motivation_
> _is extremely important in facilitating students' investment in,_
> _and progress made in, improving their writing abilities._
>
> (Vandrick, 2003: 264-265)

Therefore, literature is a sort of stimulus that encourages the learners to express their ideas and respond to the given text in a written form by utilizing of the information they read to support their opinions. The learners' motivation enhances their performance and gives them the possibility to use the writing skill as a means of communicating with the teachers and their peer students. Teaching literature to learners of English as a foreign language can improve their mastery of the English language both orally and in writing (Miliani, 2003).

In fact, literature enables the students to practise analysing and interpreting ideas by thinking and writing in a multidimensional and analytical manner. In this way, literature leads to critical thinking which is an essential feature of a good writer who must be able to question, analyse, interpret and synthesize the data provided in a literary text so that to comment on them via the writing skill.

Like any other form of readings, literature helps to learn writing as it exposes the learners to language patterns and specific vocabulary that can be stored and retrieved later on to produce essays (Vandrick, 2003).

Composition studies also rely on the use of genres. The term genre has a variety of meanings. Traditionally, it referred to the different literary types of prose and poetry. Recently, it has been employed to denote the different types of texts or discourse categories. Therefore, genre means a type of text. This type may take the form of a conversation, a letter, a poem, a recipe, a play or an essay. This term has been adapted to refer to the types of communicative events. A communicative event means an oral or written interaction containing a complete message. The event may be a dialogue, a newspaper advertisement, a novel or a shopping list. Thus, the text type is called the event by linguists. Each event implies a specific communicative purpose that is reflected in the structure of the text which embodies a particular type of discourse (Nunan, 1993).

Genre is a cognitive and social concept. It is cognitive as it involves knowledge that is linked to schemata; it is social since it enables the readers to share the same knowledge of the writer about a specific type of discourse.

Teachers of writing can develop their learners' written productions by making them familiar with genres (Johns, 2003). Each text is distinct from the other which leads it to have specific discourse characteristics. The differences in texts are classified into three categories which are generic structure, grammar and physical appearance.

Generic structure refers to the organization of ideas in a text. It often includes three steps which involve an introduction, the orientation or development and a conclusion. These steps may change from a genre to the other. For example, in a recipe, the generic structure embodies a title, a list of ingredients and a procedure while in a narrative essay, the introduction, orientation and conclusion are provided.

Grammar also differentiates between genres. For instance, the tense used in a recipe is the imperative. An article may contain boosting devices such as the adverbial constructions: 'without doubt, unquestionably and inevitably which strengthen the given statement. Certain texts may include downtoners which are adverbs that weaken the force of statements like 'slightly' and 'almost'; they can contain hedges which are epistemic certainly markers such as '...suggest that..', 'somewhat' and 'kind of'.

Thus, another grammatical feature characterizing different genres is epistemic modality which is a device used by writers to indicate the degree of commitment to the idea stated. It relies on the employment of boosting and downtoners. In fact grammatical structures help to distinguish between the various types of texts. For example, an account or a short story focuses on the past tense and may involve the use of reference, subordination and conjunction. The other distinctive characteristic of genres is the physical appearance which refers to the layout or the form of the text. For example, an article is composed of a headline, and columns including separate paragraphs; a recipe is in the form of a list followed by a paragraph. (Nunan, 1993)

Teaching the students the different genres and the distinctions that exist between them can help them to read texts and study their structures in order to produce their own pieces of writing that are in conformity with the studied texts.

The instruction of composition through genres is based on a process of needs assessment. Thus, the teachers select specific genres to be taught on the basis of their learners' needs. "*Some genres are more useful when dealing with concrete representations and are therefore fundamental in building understanding through language*" (Knapp and Watkins, 2005: 84). In this way, certain texts may be a means of explaining and describing abstract concepts and language so that to facilitate to the learners the formulation of their understanding of the text in a written form.

Johns (2003) suggests the use of various models to teach writing via genres. The first model is based on the elemental genres which have been identified by the Sydney school linguists and practitioners in the 1980's. These elemental genres are: discussions, procedures and narratives. Discussions are found in public forms, research summaries and essays. Their generic structure involves the issue, arguments and a conclusion. Procedures exist in science experiments and instructional manuals. Their schematic structure includes the goal, the steps of the procedure and the results. Narratives are found in novels and short stories. Their generic structure includes the orientation, the complication, the evaluation and the resolution.

Teaching these elemental genres to the students enables them to comprehend the social purposes of the genres and the contexts in which they can be encountered as well as the various stages of the text. Therefore, they will make a difference between genres and develop their writing level by generating a variety of essays that display a distinction in their purpose, organization and language.

The second model that can help in teaching compositions is 'the genre-based cycle of teaching and learning' which is based on the work of the Russian psychologist Vygotsky (1978) which was interpreted by Bruner (1986). This cycle relies on an interactive process that consists of five steps which are: building the context, modelling and deconstructing the text, joint construction of the text, independent construction of the text and linking related texts. *"The first step involves assisting students in understanding the context for writing"* (Johns, 2003: 203). In this stage, the teacher helps the learners to build a context by exploring the cultural aspects of the social context of the target text through exercises that focus on brainstorming, reading relevant materials, viewing realia and vocabulary activities.

The second step involves the use of a variety of texts dealing with different genres to make the students understand how the writing stages and purposes as well as language interact to generate various genres. Thus, the learners analyse and deconstruct the given texts to describe the distinct steps of the development of the topic. This phase aims at drawing their attention to the text's structure and language features.

The third step puts the emphasis on the joint construction of the text by encouraging the learners to work together. In this way, groups of students build a text corresponding to the target genre while the teacher guides them in the drafting process. The next stage is devoted to the independent construction of the text as each learner has to work alone. The students write their own texts independently. Finally, the learners have to produce compositions related to different genres (Feez, 2002). Therefore, this cycle employs a process of analysis and discussion as well as a collaborative and individual construction of texts. This technique enables the students to study how writing functions in order to give them the possibility to produce their own versions of the analysed texts.

Another procedure that is very useful in the instruction of composition is genre analysis. Johns (2003) asserts that the use of genre for teaching writing has been influenced by the most famous research in ESP genre analysis which was conducted by John Swales in the 1980's. Swales summarised his findings in his book, <u>Genre Analysis: English in Academic and Research Settings</u> (1990). He analysed academic research articles and classified the various steps of an article into functional sections referred to as the moves. A move is a basic unit in the structure of a text. The model of Swales comprises three moves. The first move is based on establishing the research field. This step is related to the introduction of the topic. The second move focuses on summarising the previous research which means that it turns around the orientation and the discussion of the topic. The third one deals with the preparation for present research by drawing a number of conclusions.

The notion of moves is employed to analyse texts as it involves the various steps that make a paper coherent to genre-experienced readers (Master, 2005). In fact, all the models of writing involve three moves which focus on introducing, developing and concluding the discussed theme. However, each move incorporates various steps that may differ from a genre to the other. The concept of moves analyses is considered as synonymous with genre analysis which is a system that describes the communicative functions that are repeated in genres and their relationship with the linguistic exponents. It combines the concept of textual awareness with the notion of how rhetorical considerations govern grammatical choices (Dudley-Evans and St. John, 1998). Genre analysis can be employed to teach the learners academic writing so that they can practise how to write essays, articles, reports, formal letters or research papers.

Raising the learners' awareness about the different genres shifts their attention from the study and analysis of texts to the production of their own pieces of writing. *"Teaching students about different genres and text forms will help them learn to organize their writing in ways that are appropriate to the topic and purpose"* (Rog and Kropp, 2004: 38). The instruction of genres explains to the

learners the nature of the interaction between the social context, the language, the purposes and the organization of the various text types. As a result, they will take into consideration the dimensions of writing including the type of the text, the purpose of writing and the audience. This degree of awareness about why and how they should write can help the students to improve their written production.

Another strategy that has a great impact on the improvement of the learners' writing performance is iterative practice which enables the students to produce a variety of essays. The learners can achieve competency in this skill through different activities.

> *They require considerable yet varied practice to develop an understanding of all aspects of writing, not only grammatical and structural knowledge of text...but also spelling, punctuation and syntax, which are vital to writing effectively. Through iterative practice students come to habituate the skills that then give them the cognitive space to reflect upon the writing process.* (Knapp and Watkins, 2005: 84)

The use of a wide range of writing activities that focus on improving the learners' linguistic knowledge and critical thinking can lead to a development in the students' writing abilities as the students manipulate the processing skills that are involved in the activity of producing compositions. To ensure the effectiveness of iterative practice, there must be a collaboration between the learners and the teacher. This kind of collaboration is seen by Vygotsky (1978) as a dialogue between the teacher and his students aiming at producing a development in the teaching and learning situation.

In fact, the success of the teaching / learning situation depends on the teacher's role which is described by Bruner (1986) as scaffolding. This means that the instructional techniques employed by the teacher facilitate learning a language in

general and the writing skill in particular. Knapp and Watkins (2005) suggest an explicit and systematic instruction of writing. Explicitness means that the students have to be taught about the nature and the various facets of this skill. Systematic instruction refers to the treatment of the different aspects of writing via appropriate stages. Thus, each taught element relates to another to enable the students to accumulate knowledge about grammar as well as writing processes through separate activities. Then, effective application takes place by requiring the learners to put into practice the skills they acquired previously in their compositions.

When teaching writing, the teacher has also to take into consideration the learners' differences as the students are classified into three groups according to the strategies they employ to write essays. These categories include planners, revisers and mixed writers. The planners focus on planning without giving much importance to the revision of their written product. The revisers rely on revision to develop the content and the structure of the essay. The mixed writers utilize extensive planning and revision in their writing (Tynjälä, 2001). Since the learners differ in their writing strategies, the identification of the distinctions in the ways how these students write will help the teacher to understand the learners' individual needs and provide them with the required scaffold for their writing. In this case, the learners' different approaches to learning writing direct the teacher's instruction of this skill.

It takes time and energy to become effective writers. Therefore, students should be helped and trained to write well. This can be achieved by motivating the learners to write essays and compositions without being afraid of making mistakes. They have to be encouraged to learn from their errors so that they will not dislike writing. According to Cimcoz (1999), the teacher should make the students understand their capacity to write and develop it.

Producing several drafts when writing an essay may help the learners to obtain an acceptable final version. In this way, the students will build a certain self-confidence since they will be convinced that their final essay is better than the initial one which

means that they can cope with their mistakes and overcome their difficulties in writing.

Also, the assessment of writing can provide diagnostic information about the learners' needs and their writing performance so that appropriate writing tasks can be designed to achieve better results. The rapid advance in technology and increased global communication in the 21st century impose new ways of perceiving the teaching of writing. Educators and teachers may be required to give more importance to the kinds of writing that are now considered as being more relevant to everyday life like electronic mail and the other forms of electronic communication (Weigle, 2002). Therefore, technology may have a great impact on the teaching and assessment of writing. However, the teachers of writing have to take into consideration the needs of their learners and the methods that will fulfil their instructional purposes.

5.15- Conclusion

This chapter has provided some suggestions and recommendations that focused on the necessity of employing a variety of techniques for developing holistic scoring and improving the testing situation. It also suggested the employment of different strategies for teaching writing. This means that writing assessment cannot function in isolation of the instruction of this skill. Both processes interact and affect the performance of the learners. Therefore, a clear perception of the objectives of teaching and testing writing is needed for the provision of suitable teaching materials and the selection of scoring procedures. However, it should be noted that these processes are always in constant change since they are related to human beings. This leads to the conclusion that any sort of suggestion is just an attempt to elicit some of the ideas that have to be discussed and applied according to the needs of the learners and the purposes of assessment.

GENERAL CONCLUSION

Testing is a means of measuring the learners' performance. It provides a description of the students' ability to use the language in order to convey a meaningful message. It also gives information about the learners' strengths and weakness concerning their educational achievement and their knowledge of certain skills or elements of language. However, tests and examinations are considered as effective and useful only if they possess four main characteristics which are: validity, reliability, discrimination and practicality.

Writing assessment involves the employment of a variety of tasks according to the purposes of testing. These tasks may include multiple choice items, cloze tests, writing sentences or essays. The rating procedure differs according to the nature of the assessment tasks. The three main scoring methods are analytic, holistic and trait-based scoring. However, writing performance assessment is often based on requiring the learners to write essays that are scored holistically.

The present research work tackled the issue of utilizing holistic scoring for assessing students' compositions. The case study involved the fourth year graduate students at the English section at the University of Tlemcen. The project was composed of five chapters. The first chapter dealt with the theoretical review of assessment and all the concepts related to it. The second chapter provided a description of the teaching and testing of the writing skill at higher education by making a special reference to the University of Tlemcen. The third and fourth chapter focused on the case study. Thus, the former chapter presented the process of data collection while the latter dealt with the analysis and interpretation of the gathered information as well as the discussion of the main research results. Finally, the fifth chapter supplied the reader with a set of suggestions and recommendations that can contribute to the development of the scoring procedures and the testing process in general.

Therefore, this research work elicited the theoretical and practical considerations related to the application of holistic scoring in the process of writing assessment. The case study provided information about different samples within the target population through the use of various research instruments.

On the basis of the collected data, one can attempt to answer the research questions that were formulated in the introductory section of the project. It can be assumed that the main issues that are related to holistic scoring turn around the degree to which this instrument is valid and reliable. In fact, the notion of the reliability of such a procedure is often discussed by researchers. Holistic scoring is not 100% reliable as certain inconsistencies are noticed when marking students' essays. This is due to the differences in the opinions and judgement of the scorers. On the other hand, holistic scoring is considered as a quite valid procedure since it helps to assess the learners' knowledge through a global impression. It reflects the learners' degree of performance which is very essential for the process of making decisions about the students.

As far as subjectivity is concerned, one can say that language testing is somewhat subjective. However, the degree of subjectivity may decrease when employing indirect methods while it increases when using direct methods. Thus, the students' response to the essay examinations involves a certain extent of subjectivity and the assessors' scoring of compositions tends to be subjective. Nevertheless, the teachers often check their subjectivity by relying on logic and a set of criteria that help in the judgement of the quality of the students' writing. This means that subjectivity can be controlled to reach a fair judgement, but it can never disappear since even the scoring rubrics or analytic scales oblige the teachers to trust their own personal judgement when deciding about each specific criterion.

In fact, the teachers face certain difficulties when correcting essay examinations. For instance, the essays of some students may display a degree of ambiguity or they may include a large number of mistakes so that the scoring

procedure becomes a heavy load. Also, the problem of form versus content is a difficult issue that is often raised especially when the responses of some learners deviate from the principal purpose of assessment. On the other hand, the students face some problems in answering essay examinations mainly because of anxiety, the difficulty of the writing skill, content factors especially the students' inability to organize ideas and language factors related to the lack of vocabulary and grammatical mistakes.

Despite these difficulties, it can be stated that some students are very successful in writing essays while the level of some other students tends to be satisfactory. The analysis of the learners' marks reveals that the learners' writing performance in the first and the second examinations is fairly average but it is moderated via the qualifying examinations to reach an acceptable level.

Generally speaking, the case study focused on the examination of the situation of writing assessment and the application of holistic scoring for testing the students' ability to produce compositions. The conclusion is that essay examinations are useful tools for assessing writing since objective tests can not reflect the learners' performance. For instance, answering multiple-choice questions does not provide information about the students' ability to express their own ideas, to discuss and to criticize certain issues.

Therefore, critical thinking and the capacity to employ the target language for communication have to be assessed holistically through direct testing methods.

The correction of essay examinations can be considered as a burden since the assessment results will have an impact on the learners. This idea of washback effect is clearly noticed at the level of the fourth year students whose future depends on their acquisition of a licence degree. However, the certification cannot be achieved without satisfactory results. This means that educational accountability has to exist in order to ascertain the fairness of the testing process and to take into consideration the kind of effects produced by the examinations. Thus, the administration and

teachers are responsible for the implementation of the instructional programme and the testing procedures as well as the learners' degree of achievement.

In fact, the assessment purposes are linked to the objectives of teaching English at university. This justifies the use of essay examinations in the majority of the taught modules. The students are prepared to sit for written examinations since their arrival to university. This kind of preparation becomes intensive from a level to the following one. The learners are gradually introduced to essay examinations so that in the fourth year, the process of testing totally relies on such examinations.

The scoring procedure involved in writing assessment is holistic as it enables the teachers to cope with a large number of copies within a limited period of time. The utilization of holistic scoring is considered as effective since it gives the possibility to provide a general view about the learners' performance on the contrary to other types of scoring which may distract the assessor and lead him to focus on one aspect rather than judging the different aspects related to the quality of the essay. As any instrument presents certain inadequacies, holistic scoring is no exception. This is why there is a sort of debate over the validity of using such a method. The main question that is raised concerns the subjectivity of the scorers which may affect the reliability of the scores.

The choice of a rating procedure has to be based on the consideration of its advantages and drawbacks as well as its suitability to the testing situation.

Holistic scoring is seen as an essential instrument in writing performance assessment. Therefore, one can suggest that the instrument is not problematic but the way it is used leads to create problems. To avoid such difficulties there must be a certain agreement on how this procedure should function. In addition to this, the elicitation of assessment criteria and rater training may help to make holistic scoring more effective.

The last chapter of this research work provided some suggestions concerning the development of writing assessment such as the use of scoring rubrics,

distributive evaluation, portfolio assessment, product analysis and the employment of the six traits for teaching and testing writing. Nevertheless, these suggestions remain mere recommendations that may facilitate the task of the assessors. They will never provide evident solutions mainly because the process of assessment and the scoring procedures are influenced by a set of factors that cannot be totally understood and controlled. This situation will always raise the question of the validity of the examinations and the instruments employed for rating them.

Therefore, the validation of holistic scoring must rely on the investigation about the degree of authenticity of the essay examinations and the washback effect produced by holistic scores. However, it is difficult to predict to what extent the learners who performed well in examinations will be able to use the acquired knowledge in real situations. Also, there is no guarantee that such examinations will always create a positive effect since the scores tend to vary from a term to the other and from an individual to the other.

The testing of social sciences in general and language testing in particular often display a certain difficulty in the interpretation of the scores which may be high or low depending on the factors that influence them. The nature of the students and raters as human beings will often reflect its subjective aspect on the process of assessment. The students may perform in a way or another due to several factors such as their degree of understanding, motivation, health condition, anxiety or apprehension of the topic. On the other hand, the scorers may assess the students' papers differently because of many reasons that may include bias, the consideration of form over content or vice versa, the teachers' experience and their personal expectations. The inconsistency of scores may be related to the quality of the students' essays and the raters' degree of appreciation of these compositions. Thus, the interference of the students' subjectivity in answering essay questions and the teachers' bias in assessing the learners' written product constitute obstacles that make it impossible to produce accurate interpretations of the assessment results.

The evaluation of the real function of holistic scoring tends to be based on how the teachers view this instrument and how the students will perceive the obtained scores. Assuming that this procedure is valid stems from the notion that it assesses what it is intended to measure. However, considering the various variables that can affect the learning and testing processes, one has to question the degree of the generalizability of the obtained results as well as the expectations of encountering similar results in different occasions.

Since the reliability of holistic scores remains problematic, the validity of holistic scoring will always raise questions that may not be totally answered. The process of validation may require a variety of research instruments as well as the investment of more time and energy but the proof of validity may remain an enigma due to the nature of the writing skill, the learners' cognitive abilities, the affective factors and the approaches employed by the teachers to teach and assess writing.

This research work aimed at providing a description of the situation of writing assessment, the different aspects that are related to the use of holistic scoring for writing students' essays and the quality of the learners' writing performance. The data collected supplied a sort of responses to the research questions and quite confirmed the stated hypotheses. Nevertheless, this project remains just a humble attempt to shed light on the direct methods of writing assessment. Further research is still needed to provide explanations to the unanswered questions and to tackle the main concerns echoed about holistic scoring and its validity.

BIBLIOGRAPHY

Bibliography

Books:

- Arora, M.G, and Vashist, S.R., <u>Theory of Educational Evaluation</u>, New Delhi: Anmol Publications, 2004.

- Arter, Judith A., and McTighe, Jay, <u>Scoring Rubrics in The Classroom</u>, UK: Corwin Press, 2000

- Babbie, Earl, <u>The Practice of Social Research,</u> USA: Cengage Learning, 2009.

- Babin Edith, and Harrison, Kimberly, <u>Contemporary Composition Studies: A Guide To Theorists and Terms,</u> USA: Greenwood Press, 1999.

- Bachman, Lyle F., and Palmer, Adrian S., <u>Language Testing in Practice: Designing and Developing Useful Language Tests</u>, Oxford: Oxford University Press, 1996.

- Bachman, Lyle F., <u>Fundamental Considerations in Language Testing</u>, Oxford: Oxford University Press, 1990.

- Bachman, Lyle F., <u>Statistical Analyses for Language Assessment</u>, Cambridge: Cambridge University Press, 2004.

- Balnaves, Mark, and Caputi, Peter, <u>Introduction to Quantitative Research Methods: An Investigative Approach,</u> London: Sage Publications Ltd, 2001.

- Banta, Trudy W., <u>Building a Scholarship of Assessment</u>, USA: John Wiley and Sons, 2002.

- Barone, Diane M., and Taylor, Joan, <u>Improving Students' Writing, K-8: From Meaning-Making to High Stakes!</u>, UK: Corwin Press, 2005.

- Benjamin, Amy, <u>Formative Assessment for English Language Arts: A Guide for Middle and High School Teachers</u>, New York: Eye on Education, 2008.

- Borton, Sonya, Frost, Alanna, and Warrington, Kate, 'Assessing The Composition Program on Our Own Terms'. in Christopher, Scheiner, (ed.), <u>Handbook of Research on Assessment Technologies, Methods and Applications in Higher Education</u>, USA: Information Science Reference, 2009, pp. 167-184.

- Bouhouche, Ammar, <u>The Researcher's Guide in Methodology and Theses Writing</u>, Algiers: ENAG Editions, 2002.

- Brindley, Geoff, 'Assessment', in Ronald, carter, and David, Nunan, (ed.), <u>The Cambridge Guide to Teaching English to Speakers of Other Languages</u>, Cambridge: Cambridge University Press, 2001, pp.137-143.

- Brindley, Geoff, 'Issues in Language Assessment', Robert B., in Kaolan, (ed.), <u>The Oxford Handbook of Applied Linguistics</u>, Oxford: Oxford University Press, 2005, pp.459 - 470.

- Brookes, Arthur and Grundy Peter, <u>Writing For Study Purposes</u>, Cambridge: Cambridge University Press, 1990.

- Brown, George, Bull, Joanna, and Pendlebury, Malcom, <u>Assessing Students Learning in Higher Education</u>, London: Routledge, 1997.

- Brown, James Dean, <u>Understanding Research in Second Language Learning</u>, Cambridge: Cambridge University Press, 1988.

- Bruner, Jerome, <u>Actual Minds, Possible Worlds</u>, USA: The President and Fellows of Harvard College, 1986.

- Buck, Gary, <u>Assessing Listening,</u> Cambridge: Cambridge University Press, 2001.

- Burton, Susan, and Steane, Peter, <u>Surviving Your Thesis</u>, London: Routledge, 2004.

- Carrol, Brendan J., <u>Testing Communicative Performance</u>, Oxford: Pergamon Press Ltd, 1980.

- Carter, Ronald, and Nunan, David, <u>The Cambridge Guide to Teaching English to Speakers of Other Languages</u>, Cambridge: Cambridge University Press, 2001.

- Celce-Murica, Marianne, Brinton, Donna,.M, and Goodwin, Janet.M, <u>Teaching Pronunciation,</u> Cambridge: Cambridge University Press, 1996.

- Chesla, Elizabeth, <u>Write Better Essays</u>, New York: Learning Express, 2000.

- Chomsky, Noam, <u>The Architecture of Language</u>, New Delhi: Oxford University Press, 2000.

- Clark, Irene L, and Bamberg, Betty, <u>Concepts in Composition: Theory And Practice In The Teaching Of Writing</u>, New Jersey: Lawrence Erlbaum Associates, 2003.

- Cline, Tony, <u>The Assessment of Special Educational Needs</u>: <u>International Perspectives</u>, London: Routledge, 1992.

- Cohen, A., <u>Assessing Language Ability in The Classroom</u>, Boston: Heinle and Heinle, 1994.

- Cohen, Louis, Manion, Lawrence, and Morrison, Keith R.B, <u>Research Methods in Education</u>, London: Routledge, 2007.

- Colburn, Alan, <u>The Lingo of Learning: 88 Education Terms Every Science Teacher Should Know</u>, USA: National Science Teachers Association, 2003.

- Connor, Ulla, <u>Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing</u>, Cambridge: Cambridge University Press, 1996.

- Corbetta, Piergiorgio, and Patrick, Bernard, <u>Social Research: Theory, Methods and Techniques,</u> London: Sage Publication Ltd, 2003.

- Corder, S. Pit, <u>Introducing Applied Linguistics</u>, England: Penguin Books Ltd, 1973.

- Cottrell, Randall R., and Mckenzie, James F., <u>Health Promotion and Education Research Methods: Using the Five - Chapter Thesis Dissertation Model</u>, Canada: Jones and Bartlett Publishers, 2005.

- Creswell, John W., <u>Research Design: Qualitative, and Mixed Method approaches</u>, London: Sage Publication Ltd, 2002.

- Crotty, Michael, <u>The Foundations of Social Research: Meaning and Perspective in The Research Process,</u> London: Sage Publications Ltd, 1998.

- Crystal, David, <u>English as a Global Language</u>, Cambridge: Cambridge University Press, 1997.

- Crystal, David, <u>The English Language</u>, London: The Penguin Group, 1988.

- Culham, Ruth, <u>6+1 Traits of Writing: The Complete Guide</u>, USA: Northwest Regional Educational Laboratory, 2003.

- Cumming, Alister and Berwick, Richard, <u>Validation in Language Testing</u>, UK: Multilingual Matters Ltd, 1996.

- Cunningham, George K., <u>Assessment in the Classroom: Constructing And Interpreting Tests</u>, Uk: Falmer Press, 1998.

- Cunningsworth, Alan, <u>Evaluating and Selecting EFL Teaching Materials</u>, London: Heinemann Educational Books Ltd, 1984.

- Dattalo, Patrick, <u>Determining Sample Size: Balancing Power, Precision and Practicality</u>, Oxford: Oxford University Press, 2008.

- Davies, Alan, and Elder, Catherine, 'Validity and Validation in Language Testing', in Eli, (ed.), Hinkel, <u>Handbook of Research in Second Language Teaching and Learning,</u>

New Jersey: Laurence Erlbaum Associates, 2005, pp.795-817, New York: Brunner-Routledge.

- Davies, Barbara Gross, Tools for Teaching, USA, John Wiley and Sons, 2009.

- Donnelly, Robert A., The Complete Idiot's Guide to Statistics, USA: Alpha Books, 2004.

- Dornyei, Zoltan, Questionnaires in Second Language Research: Construction, Administration and Processing, UK: Taylor and Francis, 2002.

- Dubin, Faida, and Olshtain, Elite, Course Design: Developing Programs and Materials for Language Learning, Cambridge: Cambridge University Press, 1986

- Dudley Evans, Tony, 'English for Specific Purposes', In Ronald, Carter, and David, Nunan, (eds.), The Cambridge Guide to Teaching English to Speakers of Other Languages, Cambridge: Cambridge University Press, 2001, pp.131-136.

- Dudley-Evans, Tony, and St. John, Jo, Developments in English for Specific Purposes: A Multi-Disciplinary Approach, Cambridge: Cambridge University Press, 1998.

- Dyer, Henry S., 'On the Assessment of academic achievement', in Richards E. Ripple (ed.), Reading in Learning and Human Abilities, New York: McGraw Hill, 1964, pp.451-461.

- Eisenhart, Margaret, 'Generalization from Qualitative Inquiry', in Kadriye, Ercikan, and Wolff-Michael, Roth, (eds.), Generalizing from Educational Research: Beyond Qualitative and Quantitative Polarization, New York: Routledge, 2008, pp.51-66.

- Elbow, Peter, Everyone Can Write: Essays Towards a Hopeful Theory of Writing and Teaching Writing, Oxford, Oxford University Press, 2000.

- Elfner, Eliot, 'Developing Rubrics as a Means for Assessment', in James O.,Nichols, And Karen W., Nichols, (eds.), A Road Map for Improvement of Student Learning and Support Services Through Assessment, New York: Agathon Press, 2005, pp. 110-114.

- Fathman, Ann. K., and Whalley, Elizabeth, 'Teacher Response to Students Writing: Focus on Form Versus Content', in Barbara, Kroll, (ed.), Second Language Writing: Research Insights for the Classroom, Cambridge: Cambridge University Press, 1990, pp. 178-190.

- Feez, Suzan, 'Heritage and Innovation in Second Language Education', in Ann M. Johns (Ed.), <u>Genre in the Classroom: Multiple Perspectives</u>, New Jersey: Lawrence Erlbaum Associates, 2002, pp. 43-72.

- Ferris, Dana, and Hedgcock, John, <u>Teaching ESL Composition: Purpose, Process and Practise</u>, New Jersey: Lawrence Erlbaum Associates, 2004.

- Finn, Seth, 'Media Writing', in William G., Christ, (ed), <u>Media Education Assessment Handbook</u>, New Jersey: Lawrence Erlbaum Associates, 1997, pp. 165 – 188.

- Finocchiaro, Mary, <u>Teaching Children Foreign Language</u>, New York: McGraw Hill, 1964.

- Fullan, Micheal, Hill, Peter, and Crévola, Carmel, <u>Breakthrough</u>, UK: Corwin Press, 2006.

- Galko, Francine D., <u>Better Writing Right Now</u>, New York: Learning Express, 2001.

- Genesee, Fred, and Upshur, John A., <u>Classroom-Based Evaluation in Second Language Education</u>, Cambridge: Cambridge University Press, 1996.

- Glenn, Cheryl, <u>The Harbrace Guide to Writing</u>, USA, Cengage Learning, 2008.

- Gomez, Stephen, and OsbornE, Richard, 'Enhancing Student Appreciation of Written Feedback on Essay Assignments', in Steve, Frankland, (ed.), <u>Enhancing Teaching and Learning Through Assessment: Deriving an Appropriate Model,</u> USA: Springer, 2008.

- Gorard, Stephen, <u>Quantitative Methods in Educational Research: The Role of Numbers Made Easy</u>, London: Continuum International Publishing Group, 2001.

- Grabe, William, 'Reading and Writing Relations: Second Language Perspectives on Research and Practice', in Barbara, Kroll, (ed.), <u>Exploring the Dynamics of Second Language Writing</u>, Cambridge: Cambridge University Press, 2003, pp. 242-262.

- Haggstrom, Margaret Austin, Morgan, Leslie Zarker, and Wieczorek, Joseph A., <u>The Foreign Language Classroom: Bringing Theory And Practice</u>, UK: Taylor And Francis, 1995.

- Haines, Catherine, <u>Assessing Students' Written Work: Marking Essays and Reports</u>, London: Routledge, 2004.

- Hamp-Lyons, Liz, 'English for Academic Purposes', in Ronald, Carter, and David, Nunan, (ed.),<u>The Cambridge Guide to Teaching English to Speakers of Other Languages,</u> Cambridge: Cambridge University Press, 2001, pp.126-130.

- Hamp-Lyons, Liz, 'Second Language Writing: Assessment Issues', in Barbara, Kroll, (ed.), Second Language Writing: Research Insights for the Classroom, Cambridge: Cambridge University Press, 1990, pp. 69-87.

- Hamp-Lyons, Liz, 'Writing Teachers as Assessors of Writing', in Barbara, Kroll, (ed.), Exploring the Dynamics of Second Language Writing, Cambridge: Cambridge University Press, 2003, pp. 162-190.

- Haney, Walter M., Madaus, George F., and Lyons, Robert, The Fractured Market for Standardized Testing, USA: Kluwer Academic Publishers, 1993.

- Hargreaves, Andy, Teaching in the Knowledge Society, England, Open University Press, 2003.

- Harmer, Jeremy, How to Teach Writing, England: Pearson Education Limited, 2004.

- Harris, David P., Testing English as a Second Language, New York: McGraw Hill, 1969.

- Harris, John, Introducing Writing, London: The Penguin Group, 1993.

- Harrison, Andrew, A Language Testing Handbook, London: Mcmillan Publishers Ltd, 1983.

- Hartley, James, Academic Writing and Publishing: A Practical Handbook, UK: Taylor and Francis, 2008

- Hawkins, Eric, Awareness of Language: An Introduction, Cambridge: Cambridge University Press, 1984

- Heaton,J.B, Writing English Language Tests, London: Longman, 1975.

- Heywood, John, Assessment in Higher Education: Student Learning, Teaching, Programmes and Instruction, London: Jessica Kingsley Publishers Ltd, 2000

- Houser, Rick, Counseling and Educational Research: Evaluation and Application, California: Sage Publications Inc., 1998.

- Hyland, Ken, Second Language Writing, Cambridge: Cambridge University Press, 2003.

- Jenkins, Sue, Price, Connie J., and Straker, leon, The Researching Therapist: A Practical Guide to Planning, Performing and Communicating Research, USA. Elsevier Health Sciences, 1998.

- Johns, Ann M., 'Genre and ESL/EFL Composition Instruction, in Barbara, Kroll, (ed.),Exploring the Dynamics of Second Language Writing, Cambridge: Cambridge University Press, 2003, pp. 195-217.

- Jones, Carys, Turner, Joan, and Street, Brian V., Students' Writing in the University: Cultural And Epistemological Issues, Amsterdam: John Benjamins Publishing, 1999.

- Kaplan, Robert B., The Oxford Handbook of Applied Linguistics, Oxford: Oxford University Press, 2002.

- Kempson, Ruth.M , Semantic Theory, Cambridge, Cambridge University Press, 1977.

- Knapp, Peter, and Watkins, Megan, Genre, Text, Grammar: Technologies for Teaching and Assessing Writing, Australia: UNSW, 2005.

- Kramsh, Claire, Context and Culture in Language Teaching, Oxford: Oxford University Press, 1993.

- Krashen, Stephen. D and Terrell, Tracy. D, The Natural Approach: Language Acquisition in the classroom, England: Pearson Education Limited, 1983.

- Kunnan, Antony John, Fairness and Validation in Language Assessment: Selected Papers From The 19th Language Testing Research Colloquium, Orlando, Florida, Cambridge: Cambridge University Press, 2000.

- Lado, Robert, Language Testing, London: Longman, 1961.

- Leech, Geoffrey, Semantics: The Study of Meaning, London: The Penguin Group, 1974.

- Lippman, Julie Neff, 'Assessing Writing', in Irene L., Clark, and Betty, Bamberg, (eds.), Concepts in Composition: Theory and Practice in the Teaching of Writing, New Jersey: Lawrence Erlbaum Associates, 2003, pp.199-240.

- Littlewood, William, Foreign and Second Language Learning, Cambridge: Cambridge University Press, 1984

- Love, Patrick, 'Document Analysis', in: Frances K., Stage, and Kathleen, Manning, (eds.), Research in The College Context: Approaches and Methods, New York: Brunner-Routledge, 2003, pp.83-96.

- Luongo-Orlando, Katherine, Authentic Assessment: Designing Performance-Based Tasks, Canada: Pembroke Publishers, 2003.

- Lynch, Brian, <u>Language Assessment and Programme Evaluation</u>, UK: Edinburgh University Press, 2003.

- Madaus, George F., Russel, Michael, and Higgins, Jennifer, <u>The Paradoxes of High Stakes Testing: How They Affect Students, Their Parents, Teachers, Principles, Schools, and Society</u>, USA: Information Age Publishing, 2009.

- Madsen, Harold S., <u>Techniques in Testing</u>, Oxford: Oxford University Press, 1983.

- Master, Peter, 'Research in English for Specific Purposes', in Eli Hinkel, <u>Handbook of Research in Second Language Teaching and Learning</u>, London: Routledge, 2005, pp.99-115.

- Mathison, Sandra, <u>Encyclopedia of Evaluation</u>, London: Sage Publication Ltd, 2004.

- Mauch, James E., and Park, Namgi, <u>Guide To the Successful Thesis and Dissertation: A Handbook for Students and Faculty</u>, USA, CRC Press, 2003.

- Maxine Earl, Lorna, <u>Assessment As Learning: Using Classroom Assessment to Maximize Student Learning</u>, UK: Corwin Press, 2003.

- McArthur, Charles A., Graham, Steve, and Fitzgerald, Jill, <u>Handbook of Writing Research</u>, New York: Guildford Press, 2006.

- McNamara, Timothy Francis, and Roever, Carsten, <u>Language Testing: The Social Dimension</u>, USA, John Wiley and Sons, 2006.

- McNamara, Timothy Francis, <u>Language Testing</u>, Oxford: Oxford University Press, 2000.

- Mertens, Donna M., <u>Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative and Mixed Methods</u>, California: Sage Publications Inc., 2005.

- Miliani, Mohamed, 'Teaching Literature in English in an EFL Situation', in Fadila, Benzaoui, and Mohamed, Miliani, (eds.), A <u>Multi-Disciplinary Approach to the Teaching of Literature</u>, Algiers: OPU, 2003, pp.21-30.

- Mitchell, Mark L., and Jolly, Janina M., <u>Research Design Explained</u>, USA: Cengage Learning, 2009.

- Moore, Tim, and Morton, Janne, 'Authenticity In The IELTS Academic Module Writing Test: A Comparative Study Of Task 2 items and University Assignments', in Lynda, Taylor, and Peter, Falvey,(eds.), <u>IELTS Collected Papers: Research In</u>

Speaking And Writing Assessment, Cambridge: Cambridge University Press, 2007,p.197-248.

- Morrow, K., 'Communicative Language Testing: Revolution or Evolution?', in C.J. Brumfit and K. Johnson (eds.), The Communicative Approach to Language Teaching, Oxford University Press, 1979, pp.143-157.

- Nation, I.S.P., Teaching ESL/EFL Reading and Writing, UK: Taylor and Francis, 2008.

- Neddar, B.A, Applied Linguistics: Schema, Discourse And Foreign Language Teaching, Oran: Edik, 2004.

- Newton, Douglas P., Teaching for Understanding, London: Routledge Falmer, 2000.

- Nichols, James O., and Nichols, Karen W., A Road Map for Improvement of Student Learning and Support Services Through Assessment, New York: Agathon Press, 2005.

- Nunan, David, Introducing Discourse Analysis, London: Penguin Books Ltd, 1993.

- Oliver, Paul, Writing Your Thesis, London: Sage Publications Ltd, 2004.

- Oskarson, M., Approaches To Self-Assessment In Foreign Language Learning, Oxford: Pergamon Press, 1978.

- Patton, Micheal Quinn, Qualitative Research and Evaluation Methods, California: Sage Publications Inc., 2002.

- Perren, G.E., 'Testing Spoken Language: Some Unsolved Problems', in Alan Davies (ed.), Language Testing Symposium: A Psycholinguistic Approach, Oxford: Oxford University Press, 1968, pp.107-116.

- Perry, Fred Lehman, Research in Applied Linguistics: Becoming A Discerning Consumer, London: Routledge, 2005.

- Phillips, Susan M., Automated Essay Scoring: A Literature Review, Canada: Society for the Advancement of Excellence in Education, 2007.

- Pilliner, A.E.G, 'Subjective and Objective Testing', in Alan, Davies, (ed.), Language Testing Symposium: A Psycholinguistic Approach, Oxford: Oxford University Press, 1968, pp.19-35.

- Popham, James, Test Better, Teach Better: The Instructional Role of Assessment, USA: ASCD, 2003.

- Porte, Graeme keith, <u>Appraising Research in Second Language Learning: A Practical Approach To Critical Analysis</u>, Amsterdam: John Benjamins Publishing Company, 2002.

- Priestley, Michael, <u>Performance Assessment in Education and Training</u>, USA, Educational Technology Publications, 1982.

- Purpura, James.E, <u>Assessing Grammar,</u> UK: Cambridge University Press, 2004.

- Race, Phil, <u>The Lecturer's Toolkit: A Practical Guide to Learning, Teaching and Assessment</u>, London: Kogan Press Limited, 1998.

- Recco, William and Schultz, Colleen, <u>501Writing Prompts</u>, New York: Learning Express, 2003.

- Richards, Jack C., and Renandya, Willy A., <u>Methodology in Language Teaching: An Anthology Of Current Practice</u>, Cambridge: Cambridge University Press, 2002.

- Richards, Jack C., <u>The Language Teaching Matrix</u>, Cambridge: Cambridge University Press, 1990.

- Roach, John, <u>Public Examinations in England, 1850-1900</u>, Cambridge: Cambridge University Press, 1971.

- Rog, Lori Jamison, and Kropp, Paul, <u>The Write Genre: Classroom Activities and Mini Lessons that Promote Writing with Clarity, Style and Flashes of Brilliance,</u> Canada: Pembroke Publishers, 2004.

- Rosano, Susan A., Gates, David M., Zaretsky, Lindy, and Kurilko, Mary Beth, 'Accountability versus Responsibility', in Joan Poliner, Shapiro, and Jacqueline Anne, Stefkovich, (eds), <u>Ethical Leadership and Decision Making in Education: Applying Theoretical Perspectives to Complex Dilemmas</u>, New Jersey: Lawrence Erlbaym Associates, 2005, pp. 120 – 136.

- Rossner, Richard, and Bolitho, Rod, <u>Currents of Change In English Language Teaching</u>, Oxford: Oxford University Press, 1990.

- Rowntree, Derek, <u>Assessing Students: How Shall We Know Them?,</u> UK:Taylor and Francis, 1987.

- Rupp, André A., And Köller, Olaf, <u>Developing Standards-Based Assessment Tasks For English as a First Foreign Language: Context, Processes And Outcomes in Germany,</u> Germany: Waxmann Verlag, 2008

- Ruth, Leo, and Murphy, Sandra, <u>Designing Writing Tasks for the Assessment of Writing</u>, New Jersey: Ablex Publishing Corporation, 1988.

- Sadoski, Mark, and Paivio, Allan, <u>Imagery and Text: A Dual Coding Theory of Reading and Writing</u>, New Jersey: Lawrence Erlbaum Associates, 2001.

- Sakyi, Alfred Appiah, 'Validation of Holistic Scoring for EFL Writing Assessment: How Raters Evaluate Compositions', in Antony John, Kunnan, (ed.), <u>Fairness and Validation in Language Assessment</u>, Cambridge: Cambridge University Press, 2000, pp.129-152.

- Salaberry, Rafael, and Cohen, Andrew D., 'Testing Spanish', in Rafael, Salaberry, And Barbara A., Lafford, (eds.), <u>The Art of Teaching Spanish: Second Language Acquisition from Research to Praxis</u>, Washington: Georgetown University Press, 2006, pp.149-172

- Sapsford, Roger, and Jupp, Victor, <u>Data Collection and Analysis</u>, London: Sage, 2006.

- Scott, David, and Morrison, Marlene, <u>Key Ideas in Educational Research</u>, London: Continuum International Publishing Group, 2006.

- Shaw, Stuart D., and Weir, Cyril J., <u>Examining Writing: Research and Practise in Assessing Second Language Writing</u>, Cambridge: Cambridge University Press, 2007.

- Silverman, David, <u>Doing Qualitative Research: A Practical Handbook</u>, London: Sage Publications Ltd, 2000.

- Speck, Bruce W., <u>Grading Student Writing: An Annotated Bibliography</u>, USA: Information Age Publishing, 1998.

- Spector, Michael, <u>Handbook of Research on Educational Communities and Technology</u>, UK: Taylor and Francis, 2007.

- Spolsky, Bernard, <u>Sociolinguistics</u>, Oxford: Oxford University Press, 1998

- Stark, Rachael, <u>Research and Writing Skills Success</u>, New York: Learning Express, 2003.

- Starkey, Lauren, <u>How To Write Great Essays</u>, New York: Learning Express, 1962.

- Stiggins, Richard J., Erter, Judith A., Chappuis, Jan, and Chappuis, Stephen, <u>Classroom Assessment for Student Learning: Doing it Right-Using it Well</u>, USA: Assessment Training Institute, 2004.

- Stufflebeam, Daniel L., Madaus, George F, and Kellaghanm, Thomas, <u>Evaluation Models: Viewpoints on Educational and Human Services Evaluation</u>, USA: Springer, 2000.

- Suskie, Linda, and Banta, Trudy W., <u>Assessing Student Learning: A Common Sense Guide</u>, USA: John Wiley and Sons, 2009.

- Sutton, Richard S., and Barto, Andrew G., <u>Reinforcement Learning: An Introduction</u>, USA: The MIT Press, 1998.

- Thissen, David, and Wainer, Howard, <u>Test Scoring</u>, New Jersey: Lawrence Erlbaum Associates, 2001.

- Thomas, Robert Murray, <u>Blending Qualitative and Quantitative Research Methods in Theses and Dissertations</u>, UK: Corwin Press, 2003.

- Thomas, Robert Murray, <u>Conducting Educational Research: A Comparative View</u>, London: Greenwood Publishing Group, 1998.

- Thornbury, scott, <u>How To Teach Speaking</u>, England: Pearson Education Limited, 2005.

- Tynjälä, Päivi, 'Writing, Learning and the Development of Expertise in Higher Education', in Päivi, Tynjälä, Lucia, Mason, and Kirsti, Lonka, (eds.), <u>Writing as a Research tool: Integrating Theory and Practice</u>, The Netherlands: Kluwer Academic Publishers, 2001, pp.37-56.

- Underhill, Nic, <u>Testing Spoken Language</u>, Cambridge: Cambridge University Press, 1987.

- Urquhart, Vicki, and McIver, Monette, <u>Teaching Writing in The Content Areas</u>, USA: ASCD, 2005.

- Van Blerkom, Malcom L., <u>Measurement and Statistics for Teachers</u>, New York: Routledge, 2008.

- Vandrick, Stephanie, 'Literature in the Teaching of Second Language Composition', in Barbara, Kroll, (ed.), <u>Exploring the Dynamics of Second Language Writing</u>, Cambridge: Cambridge University Press, 2003, pp.263-284.

- Vygotsky, Lev Semenovick, <u>Mind in Society: The Development of Higher Psychological Processes,</u> USA: The President and Fellows of Harvard College, 1978.

- Wallace, Michael J., <u>Training Foreign Language Teachers</u>, Cambridge: Cambridge University Press, 1991.

- Wallen, Norman E., and Fraenkel, Educational Research: <u>A Guide to the Process</u>, London: Routledge, 2001.

- Weigle, Sara Cushing, <u>Assessing Writing</u>, Cambridge: Cambridge University Press, 2002.

- Weir, C., <u>Understanding and Developing Language Tests</u>, New York: Prentice Hall, 1993.

- Welch, Catherine, 'Item and Prompt Development in Performance Testing', in Steven M., Downing, and Thomas M., Haladayna, (eds.), <u>Handbook of Test Development</u>, New Jersey: Lawrence Erlbaum Associates, 2006, pp.303-328.

- Wellington, Jerry, <u>Educational Research: Contemporary Issues and Practical Approaches</u>, London : Continuum International Publishing Group, 2000.

- Whiston, Susan C., <u>Principles and Application of Assessment in Counseling</u>, USA: Cengage Learning, 2008.

- Whithaus, Carl, <u>Teaching and Evaluating Writing in The Age of Computers and High-Stakes Testing</u>, New Jersey: Lawrence Erlbaum Associates, 2005.

- Wiggins, Grant P., and Mctighe, Jay, <u>Understanding by Design</u>, USA: ASCD, 2005.

- Wilkinson, David, and Birmingham, Peter, <u>Using Research Instruments: A Guide for Researchers</u>, London: Routledge, 2003.

- Williams, Marion, and Burden, Robert L., <u>Psychology for Language Teachers</u>, Cambridge: Cambridge University Press, 1997.

- Wilson, John A.R., Robeck, Mildred C., and Michael, William, <u>Psychological Foundations of Learning and Teaching</u>, New York: McGraw Hill, 1969.

Pamphlets:

- <u>The University Guide</u>. Tlemcen: The Local Education Authority of The Planning of Counselling and Information, 1991.

- <u>The Rules of the Pedagogical Organization and Management: Methods of Evaluation and Progression.</u> Tlemcen: University of Abou Bakr Belkaid, 2009.

- <u>The Student's Internal Regulations.</u> Tlemcen: University of Abou Bakr Belkaid.

- <u>Curriculum</u>. Tlemcen: Department of Foreign Languages.

Theses:

- Belmekki, Amine, 'The Importance of Text Selection and Metacognitive Awareness Raising in EFL Reading Strategies: The Case of First Year Students, Tlemcen: Abou Bakr Belkaid University- Tlemcen. Unpublished Doctorate Thesis. University of Abou Bakr Belkaid, Tlemcen, 2007.

- Benmansour – Benyelles, Radia, 'Teaching ESP at Post Graduation Level: The Case of First Year Post Graduate Student of Economics at Abou Bakr Belkaid University- Tlemcen'. Unpublished Doctorate Thesis, University of Abou Bakr Belkaid, Tlemcen, 2008.

- Benmoussat, Smail, 'Mediating Language and Culture: An investigative analysis of the Cultural Dimension in the Algerian newly designed ELT Textbooks'. Unpublished Doctorate Thesis. University of Abou Bakr Belkaid, Tlemcen, 2002.

- Hamzaoui – El Achachi, Hafida, 'An Exploration into the Strategies Used for Essay Writing Across Three Languages: The Case of EFL University Students'. Unpublished Doctorate Thesis. University of Abou Bakr Belkaid, Tlemcen, 2005.

- Ourghi, Rachid, 'Implications of Critical Reflection and Discourse Awareness for Educational Change: The Case of Writing Curriculum, Learner and Teacher Development at the University Level'. Unpublished Doctorate Thesis. University of Abou Bakr Belkaid, Tlemcen, 2001.

WEBOGRAPHY

Webography

- Albertini, John, 'Classroom Assessment of Writing: Purpose, Issues and Strategies' in: http://www.pepnet.org/confpast/1996/pdf/albertini.pdf.

- Bachor, Dan, 'Reformating Reporting Methods for case studies, Sydney: The Australian Association for Research in Education, 2000' in: http://www.aare.edu.au/oopap/bac00287.htm.

- Beattie, kate, and james, Richard, 'Assessing Essays' in: www.cshe.unimelb.edu.au/pdfs/assessing_essays.pdf.

- Cimcoz, Yesim, 'Teaching ESL/EFL Students to Write Better'. In The Internet TESL Journal, Vol.5, N° 10, (1999), in: http://iteslj.org/Techniques/Cimcoz-writinghtml.

- Clapham, Caroline, 'Principles of Assessment', UK: subject centre for languages, 2003. In: http://www.llas.ac.uk/resources/gpg/1398.

- Davis, Barbara Gross, 'Tools for Teaching', USA: Jossey - Bass, 1999, in: http://honolulu.hawaii.edu/intranet/committees/FacDevCom/guidebk/teatchtip/quizzes/.htm.

- Elander, James, 'Student Assessment From a Psychological Perspective'. In Psychology Learning and Teaching, volume 3, number 2, (2003) pages 34-41. / in: http://www.writenow.ac.uk/assessmentplus/documents/Elander%202004%20PLAT.pdf

- Harrington, Katherine, Elander, James, Norton, Lin and Robinson, Hannah, 'Do Essay Assessment Criteria Refer to Transferable Skills, Deep Approaches to Learning, or Complex Learning?', in: Investigations in University Teaching and Learning, volume1, number 2, (2003), in http://www.assessmentplus.net.

- Jacobs, Lucy. C, 'How To Write Better Tests: A Handbook For Improving Test Construction Skills', Indiana University: Bloomington Evaluation Services and Testing, 2004. In: http://www.indianna.edu/-best/write_better_.html.

- Kfouri, Carol Ann, place: Translation journal Goff 'Testing And Evaluation in The Translation Classroom' in http: //accurapid.com/journal

- Marnell, Geoffrey, 'Measuring readability', in: www.abelard.com.au/readability%20statistics.pdf.

- Morgan, Chris , 'Selecting Methods of Assessment', Oxford University, 1999, in http://www.brookes.ac.uk/services/ocsd/2_learntch/methodes.html.

- Mowl, Graham, 'Innovative Student Assessment', 1996, in:
http://www.londonmet.ac.uk/deliberations/assessment/mowl.cfm.
- Nott, David, 'Marking Students' Written Works: Principles and Practice' UK: subject
centre for languages, 2008, in: http://www.llas.ac.uk/resources/gpg/2956.

APPENDICES

Appendix A

The Questionnaires

TEACHERS' QUESTIONNAIRE

This questionnaire aims at collecting data concerning teaching and testing English at the university level. Thus , I would be grateful to you if you answer the following questions .
Thank you for your help and collaboration

1- *What are the modules you teach ?*

… …

2- *What are the difficulties you encounter when teaching English ?*

… …

… …

3- *When teaching any module , your main purpose is to make your students master :*
(provide a percentage)

❑ *The English language* … … … … … %

❑ *A specific knowledge related to the taught module* … … … … … %

4- *What kinds of questions do you use in examinations ?*

❑ *Questions requiring short answers*

❑ *Questions requiring an essay*

❑ *A combination of questions involving short answers and others requiring a paragraph*

5- *What are the difficulties you find when correcting examinations ?*

… .

… ..

… …

6- *When correcting the examination papers , you are more interested in :*

❑ *Form*

❑ *Content*

❑ *Form and content*

7- *How many points (out of twenty) do you assign to form and content :*

- *Form* ⟶ … … … …

- *Content* ⟶ … … … …

8- *Do you find it easy to prepare the questions of the examinations ?*

❑ *Yes* *why ?* …

… …

❑ *No* *why ?* …

… …

9- *What kind of mistakes do most of the students make when answering the examination questions ?*

… …

… …

… …

10- *What is your opinion about your students' level ?*

❑ *Weak*

❑ *Average*

❑ *Good*

STUDENTS' QUESTIONNAIRE

The purpose of this questionnaire is to gather information about learning English at the university level. Thus , you are kindly requested to answer the following questions .

Thank you for your collaboration

1- What are the difficulties you find in learning English ?
....
....
....

2- Which modules do you find easy to learn ?
....
....

3- What are the modules you consider as being difficult or requiring much effort ?
....
....

4- Do you try to search for more information in addition to that provided by the lectures ?
❑ *Yes*   *If yes , how ?* ....
                ....
❑ *No*   *If no , why ?* ....
                ....

5- Are you influenced by your mother tongue when learning English ?
❑ *yes*   *If yes , how ?* ....
                ....
❑ *No*

6- When required to sit for an examination , you feel :
❑ *Anxious*
❑ *Afraid*
❑ *At ease*

7- What are the difficulties you face when answering the questions of the examinations ?
....
....
....

8- Do you like the examinations that require from you to write an essay ?
❑ *Yes*   *why ?* ....
                ....
❑ *No*   *why ?* ....
                ....

9- When answering the question of any examination that requires an essay, you prefer to :
❑ *Use your own style*
❑ *To retrieve the lecture*
❑ *To combine your style with the one used in the lessons*

Appendix B

The Questions
of the Interviews

The purpose of this interview is to collect data about the teaching of English at university as well as the assessment of the writing skill. You are kindly requested to answer the following questions:

1- What is the content of the syllabus of each module you teach?
2- What are the types of exercises you often employ to consolidate the lectures?
3- What are the kinds of questions you utilize in the examinations?
4- Which methods do you use to asses the learners' responses?
5- Which methods do you employ to assess the students' dissertations?
6- What are the difficulties you find in evaluating students' projects?
7- What is your opinion concerning the learners' ability to write memoirs?
8- What are the main errors produced by the learners in their dissertations?

Thank you for your collaboration

Appendix C

The Tests' Questions

- "Colonization had serious effects on the colonized countries." Discuss

- "The novel is history as it might have happened." Discuss.

- "Language is a systematic means of communicating ideas and feelings by the use of conventionalized signs." Discuss.

Appendix D

A sample of the essays
produced by the students
responding to the tests

"Colonization had serious effects on the colonized countries." Discuss

when we look towards the word "colonization", what it me
we can say that is: one country enter to the other country, and t
to enforced its rules and system of governing in such domain
so all as, economicly and politically, this country may be
colonized other country because of the latter one raw mater
it strategy in the world or other interests which made the
other one more interested and well-prepared to colonize.

So, in fact that this political movement had many serie
effects on the colonies, fo exemple the "language" had much
affected, this effect we espeacially observe it in the dialects
~~where in some words, when we~~ th as we found in our dialects
where we mixed between French and our mother tongue Arabi
other side colonies might affected a bi's developement; soca
economicls as what happened in Algeria during the French
colonization; the Algerian society was very simple, rural
and people knew simple works which made our economic
very lazy but After the colonization, it developed in ways of
living, it changed from rural to urban one, people knew
new industries.

There are other effects, which were affected by
colonies, either in positive side or in negative one.

" Colonization had serious effects on the colonized countries." Discuss

In the late 18th century and early 19th's colonization was a fashion because of the industrial revolution. It had many effects on the colonized countries.

On one hand we find social effects such as poverty. Colonies mainly the non-developed countries suffered from unemployment and starvation and agrate loss in population and many other problems.

On the other hand the colonized countries' raw materials were used in the industries of the colonizers without any profits for their owners. The population was used for hard works and was considered as column consumers who buy goods at any price and without paying attantion to the quality.

The influence of language can be also mentioned for example im Egyptian dialect is influenced by English and Algerian dialect contains many French words

As a conclusion, the effects of colonialism are various and can't be limited and some of them are still existing.

" The novel is history as it might have happened." Discuss

The novel is history as it might have happened, when the novelist narrate real historical event who like a story about a person who was against the colonization by practicing in the war and realized a lot of success so he became a hero or thanks to him & his country became independence so, this person became a hero but not only in the novel but also in the reality because everyone know him, in this aspect the novelist only repeated what happened in the past. But sometimes the novelist speak about a hero who was a good person he, has a strange effort and he face an evil forces in last he success in this aspect all these events was only an imagination of the novelist who have is aim to show that the good nature could always success. except if these events happened with a prophets like americals.

"The novel is history as it might have happened." Discuss

Novel is a prose with an acceptable length, where the writer can expose his or other's own experiences. It is most of the times written to describe a social problem and others times find solutions to such problems.

Novel appeared clearly in the 18th because of some problems in the society, like inequality between classes like aristocracy and workers class, as there were many writers were fond of persons during a specific moment like: "Pamela" of Samuel Richardson, in addition to this, Tome Jones, Clarissa ... all these characters presented figures within society especially the British society in 18th. Pamela described how the marriage could change any person way of life, for her, she became a rich after the poverty, for Andrew Joseph, the man who protected his clarissity even if he is a man. On the other hand, Novel, ... though Novel people can describe a period of time, using the objectif critic, for example, the 18th c was described as the age of deseases, since the rich become richer and poors becoms poorer, as there are some of writers Novelists who described people as animals [dehumanousthe]. So, the Novel was as a mirror to the social, economic and politic life of the age that it appeared the 18th, and thanks to some novels, who we knew about that age, and then we may conclude that "This the mirror of the history". Regarding Novel, most of the novels were considered as an expose of the historical events: Cristopher Columbus and his discovery of America, Napolean Bonapart and his wars in Europe & America..., and others novels were written because of some political events, like a death of a president and an election results.

So, Novel, as a litterary work, is a mirror of the society, as it is an expose of the historical events. And may be these events lead to the Novel appearance, since writers were looking for a way to criticize the bad living condition in both fields society's economy.

" Language is a systematic means of communicating ideas and feelings by the use of conventionalized signs." Discuss

Language is a system of arbitrary vocal symbols by which the members of any society intract interms of cultural domains.

Language is that organized system consisted of many fields, such as phonetics which deals with the study of speech sounds, grammar interms of morphology (the form of the word), and syntax (how these words are combined to give meaningful sentence In addition to semmantics which studies the mean of words in a sentence. Each field of these has a specific aim so that no field could interfeare in another one but this doesn't mean that each down works alone; moreover, this system of communication between members of the speech community will be easier by using some conventionalized sound interms of speaking, and other conventionalized letters, interms of writing.

As a conclusion, I may mention that language is a purely human and non-instictive method of communicating ideas, feelings and emotions emotions by members of a specific speech commun

" Language is a systematic means of communicating ideas and
feelings by the use of conventionalized signs." Discuss

Language has many definitions. We can say that it's
an important means of communication. We are need the language
in our daily life. In this way, there are two famous forme
who describe lge differently. For example "De Saussure"
distinguished "lge" as "competence" and "performance", where,
"Chomsky" studied "lge" as "langue" and "Parole".
We can express language by spelling and spoking forms
using different words and sentences to have a meaning
and make the audience understand. We can also use
the jestures (and (in writing) sings and symbols.

"Language" has many advantages in our life, It helps us to
express what we want to telling, it use also to communicate such
as in TV, Telephone, Radio ..., we can use "lge" in writing to
make, sentences ....

As a conclusion, we can not give the exact difinition for "lge"
It has a hole meaning but it is good means of communication
any one can he uses it in his daily life.

Appendix E

A sample of students' examination papers

The questions:

1. State and explain four "4" by languages which derive from "Latin".

2. State and explain the three "3" domains which linguists have focused their Research on in Competency based Approach.

The Answers:

1. The four "4" languages that derive from Latin are: French, Italic, Portuguese, Spanish. These languages belong to the Indo-European languages, and these languages are in the head of the Italic language.

2. Linguists have focused their Research on: (in the Competency based Approach)

1) The Lge Learning Aptitude: Is the person's Native intellectual capacity to learn a lge. It is related to the person general level of cognitive development or Maturity.

2) Attributes and Motivation factors. They had do with learners Attitudes towards lge learning in general and the language to be learned in particular and also his Motivation to learn.

3) Personality characteristics. They include extroversion, sense of humour, lack of inhibition, Assertiveness (authority), risk taking, flexibility, confidence, determination.

II-

This literary passage is taken from a novella « Heart of Darkness » which is written by « Joseph Conrad ». This story is a modernist one and it is a metaphysical voyage in the human psyche and discovery the of the forces within the human heart.

It takes place in Africa and it is the inner station

time? The characterization is « Marlow » who is the narrator in this passage. incomplete

He describes « his memory was like the other memories of the dead » it is a symbol of death.

« He lived then before me » it is also a symbol of death?

« A shadow darker than the shadow of the night » it is a symbol of darkness « subconsious ».

« The gloom of the forests, the beat of the drums .... darkness. It is a symbol when the man is stricken to death by gloom brooding over a crowd of men. He describes it as the regularity of heart's beat of a conquering darkness.

We can say that teaching and learning are related with each other, so teaching is facilitating learning and a guiding and facilitating learning whereas learning is an acquiring of getting a knowledge of a subject or skills through different methods which are the Silent Method, the Grammar translation Method and the Audio lingual Method and the Communicative Method, so these are several ways and Technics that teaching and learning are performed in the classroom which are Styles and Strategies. If we begin by the term Style which refers to the consistant and rather enduring tendencies or preferences within an individual styles are those general characterestics of intellectual functioning and personality type as well) that especial pertain to you as an individual that differentiate you from someone else for example you might be more visually oriented, more tolerant of ambiguity o more reflective than someone else, these would be styles that characteize the General pattern in our thinking or feelings this is (in one hand, on the other hand, Strategies are a specific methods of approaching a problem or a talk modes of operation for achieving a particular end, planned designs for controlli and manipulating certain information, these are context contextualized (Battle plans) that might vary from moment to moment or day to day or year to year.

By the 19th century, the African continent witnessed many european (invisions), which were aiming in reality to colonize this continent.

Europeans exploited the musimum they can from the Africans, since they were attracted by the African's raw-materials and various resources. each nation (France, Germany, Britain) wanted to take more territories so they used whatever means they got of and other by sending missionaries, sign dafts with the chiefs of the tribes, inorder to buy lands, in addition to the the Berlin Conference played a great role in making the partition of Africa legal between these nations and which at the same time opened the doors to the European expansions in Africa. As a result Africans people did not share the same view or reaction towards colonists, some of them (1st) accepted the Europeans colonialism because they thought that fighting was unless, in addition to that the chiefs of the tribes wanted to pressure their position in the tribes. The second catigory was closure to negotiate because they considered Europeans as being powerful like what happen to some tribes yorouba they were fear to defeat by Europeans, when they saw the same bad situation of other tribes. Unlike Yorouba, Botswana's chiefs signed dafts with Europeans (he was fear from loosing the throwen). The 3rd category chose to resist was more powerful and chose to resist but it was defeated by Europeans (colonists) because of many reasons, there was lack of unity, lack of organization and because of the ignorance of those natives in using weapons and of cause lack of weapons quantities (quantities).

Generally, the Africans tried to get their freedom, but their resisting, but their aim was broken because of many reasons, which allow the European colonists to last for a long period of time in Africa.

ملخص :

تقييم الطلبة الدارسين للغة الانجليزية بالجامعة الجزائرية يعتمد على استخدام وسائل متعددة. التقييم الشامل يشكل الطريقة الأكثر استعمالا لتصحيح اختبارات الطلبة لكن هذه التقنية يمكن أن تتسم ببعض الإشكاليات التي تستطيع أن تعيق عمل الأساتذة عند تقييمهم لمستوى الطلبة.

هذا البحث يهدف إلى جمع معلومات عن طريقة التنقيط الشمولية عند تقييم المهارات الكتابية للطالب استنادا إلى حالة طلبة السنة الرابعة جامعي بجامعة أبو بكر بلقايد بتلمسان، مع تسليط الضوء على نقاط الضعف و القوة في هذه الطريقة المتبعة و كذا الصعوبات التي يواجهها الأستاذ و الطالب في مرحلة التقييم.

الكلمات المفتاحية : التقييم – التنقيط الشمولي – مهارة الكتابة – التعبير الكتابي – الاختبار.

Résumé :

De diverses méthodes sont utilisées pour examiner les étudiants de l'Anglais à l'université Algérienne. Le marquage holistique semble être le procédé le plus commun qui est impliqué dans l'évaluation de l'écriture. Cependant, cette technique peut montrer certaines issues problématiques pour les professeurs en marquant les essais produits durant les examens. Ce travail de recherche fournit des informations au sujet de l'utilisation du marquage holistique pour l'évaluation des performances de l'écrit des étudiants prenant comme référence le cas des étudiants de la quatrième année à l'université d'Abou Bakr Belkaid à Tlemcen. Ce travail essaye de clarifier les points de force et de faiblesse de ce procédé aussi bien que les difficultés rencontrées par les étudiants et les professeurs durant le processus d'évaluation.

Mots clés : évaluation, marquage holistique, performance de l'écrit, essai, examen.

Summary :

Various methods are employed for testing the learners of English at the Algerian university. Holistic scoring seems to be the most common procedure that is involved in the assessment of writing. However, this technique may display certain problematic issues that may create difficulties to the teachers when scoring essay examinations. This research work provides information about the use of holistic scoring for the learners' writing performance taking as a reference the case of fourth year graduate students at Abou Bakr Belkaid university at Tlemcen. It attempts to elicit the strengths and weaknesses of this procedure as well as the difficulties encountered by the students and the teachers in the testing process.

Key words: testing, assessment, holistic scoring, writing performance, essay, examination.