The People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed Kheider University of Biskra

Faculty of Letters and Foreign Languages

Department of Foreign Languages

Section of English

Thesis submitted to the Department of Foreign Languages in Candidacy for the Degree of Doctorate in Language and Civilization

The Washback Effect of an Alternative Testing Model on Teaching and Learning:

An Exploratory Study on EFL Secondary Classes in Biskra.

Candidate: **HOADJLI Ahmed Chaouki**　　　Supervisor: **Dr. HAMLAOUI Naima**

Board of Examiners:

Chair: Pr. HAMADA Hacène - ENS Constantine.

Supervisor: Dr. HAMLAOUI Naima - University Badji Mokhtal, Annaba.

Examiner: Pr. MELLOUK Mohamed - University Djilali Liabes, Sidi-Bel-Abbes.

Examiner: Pr. MERBOUH Zouaoui - University Djilali Liabes, Sidi-Bel-Abbes.

Examiner: Dr. BOUHANIA Bachir - University Ahmed Draia, Adrar.

March 2015

Dedication

They say it takes a minute to find special people,

an hour to appreciate them, a day to love them,

but then an entire life to forget them.

Take the time … to live and love…

(unknown author)

This doctoral thesis is dedicated to:

*My father **Athman**, My mother **Djamila***

*My wife **Samia**,*

*My daughter **Soundouce**.*

*My **brothers** and **sisters***

*My father in-law **Hadj Kaddour Moussa***

*My **brothers** and **sisters in-law***

My thanks for their prayers and support.

Acknowledgements

Abstract

This study sets out to explore the issue of evaluation in general and language testing in particular in the Algerian educational system. In precise terms, the present research investigated how a particular type of language tests, achievement tests, should be designed and developed in order to make them fit their intended purposes. This aimed to provide EFL teachers at the secondary school level in Biskra region with theoretical and practical understanding, with the expectations to make these developed tests become part of the learning experience and not a mere operation of grading for the sake of passing or failing. The rationale for this study is twofold. Firstly, given that there is a multitude of washback studies of language tests on teaching and learning, no empirical study of this sort was available for this particular context. Secondly, this study stems from the students' low scores that do not actually reflect their performance by virtue of the non-pertinence between what these students learn and what they are tested in. Ultimately, the fundamental objective of the present exploration is to yield practical insights to overcome the array of anomalies present in the current testing system. In order to achieve this objective, an Alternative Testing Model was proposed. This model was first tried-out and then put into practice on a selected sample of EFL classes in Biskra region. In the investigation of the washback effects of the new testing system on participants' perceptions of, and attitudes towards, teaching and learning, a qualitative and quantitative study was carried out, employing four data collection methods to gather insightful feedback. From the analysis of these data, the findings show that the new testing policy had positive washback effects on teachers and students in this study. It succeeded to change to some extent and with different degrees the participants' attitudes and behaviours in teaching, learning, and assessment practices.

Key Terms: Washback - Achievement Test - Test Design - Test Development - ATM

List of Abbreviations and Acronyms

ATM : Alternative Testing Model

Bac : Baccalaureate examination

COLT : The Communicative Orientation of Language Teaching

EFL : English as a Foreign Language

ELT : English Language Teaching

GTA : Grammar Translation Approach

HKCEE : Hong Kong Certificate Education Examination

MET : Matriculation English Test

NUE : New Use of English

RP : Received Pronunciation

RUE : The Revise Use of English

SPSS: Statistical Package for Social Sciences

SQ$_1$: Student's Questionnaire in the Preliminary Study

SQ$_2$: Student's Questionnaire in the Final Study

TOEFL : Test of English as a Foreign Language

TQ$_1$: Teacher's Questionnaire in the Preliminary Study

TQ$_2$: Teacher's Questionnaire in the Final Study

List of Appendices

Appendix 1: The consent letter

Appendix 2: Questionnaire validation form

Appendix 3: Teacher's questionnaire (01)

Appendix 4: Student's questionnaire (01)

Appendix 5: Bac exam guide

Appendix 6 : Sample of achievement tests based on the current testing system

Appendix 7: Sample of achievement tests based on the ATM

Appendix 8: Teacher's questionnaire (02)

Appendix 9: Student's questionnaire (02)

Appendix 10: Interview with teachers

Appendix 11: Interview and focus groups with students

List of Tables

List of Figures

Contents

XIII

General Introduction

It may seem paradoxical but tests are not sanctions. They should be looked at as a rewarding experience. In the past, language tests used to be regarded as the students' "bête noire". Students did not enjoy taking tests and teachers did not enjoy marking them. Nowadays, more focus is put on the relevance of language tests to the teaching operation. Instead of being a separate subject that frequently takes place at the end of a course, a term or an academic year, language tests have become an integral part of the curriculum; that is, they are seen as a learning experience, which is part of the on-going course of study.

More importantly, language tests are now regarded as a valuable tool for providing information that is relevant to several concerns in language teaching. They can be one way of providing systematic feedback for both teachers and students. The teacher can see how well or badly the students are performing and check for any discrepancies between expectations and actual performance. Likewise, the students can know how much attainment and progress they are doing in learning the language.

Language tests can also be a good means of evaluating instructional materials and tasks and their relevance to the educational goals. Ideally, the goals of language tests and/or test-items should be clear to students, so that they need not spend time guessing what the teacher means. If the students perceive the tests as relevant to their needs, they themselves are probably going to engage more actively in the process of dealing with them.

Another aspect of language tests concerns the insights and inferences a teacher often draws from the outcomes of language tests. The usefulness of such inferences is manifested when they provide feedback to be utilized in making the teaching programme more effective and when they provide information to the kind of materials and tasks students need. These inferences can be the only ground on which teachers can make appropriate decisions to the teaching operation.

To be able to achieve its aims, a language test must meet the requirements of some fundamental test-qualities, such as, validity, reliability, and practicality. In other contexts, it should be authentic, interactive, and should have an impact on all the concerned parties. Certainly, these considerations will vary from one situation to another because what might be appropriate in a given case might not be so in another. Because of this reality, an understanding of how one designs and develops a language test is very crucial.

Furthermore, an important assumption behind the numerous uses of language tests is that a language test exerts an influence on both teaching and learning. It can transgress these levels to concern even border areas in the educational setting. In the assessment literature, there is now a clear consensus that the impact of a test on teachers and students is termed washback. For many specialists, the concept of washback direction encapsulates the principles that some effects of a language test may be beneficial, while some other effects may be harmful. A positive washback is often seen as to encourage learning, and, conversely, a negative washback usually inhibits the attainment of the educational goals held by learners. In this sense, washback is judged positive or negative according to how far it enhances or reduces forms of teaching and learning evaluated to be appropriate. Of course, what is considered appropriate will depend on the instructional goals one espouses.

Enchanted by the power of the decisions they can provide, language tests are key components in projects to introduce in education through interventions. It is argued that a language test can stand as an effective instrument to reform or innovate an educational or a language testing system. In line with this assumption, research on washback has demonstrated that a language test can also be a way by which teachers' and students' behaviours and perceptions of their own abilities can change. In addition, a testing instrument may influence the content and methodology of teaching programmes, attitudes towards the value of certain educational goals and activities; in the long term, it many even serve the

needs of society as a whole. Given this potential power, a language test, as an innovative act, can thus become a useful source for contending participants who seek to bring about a new vision of what education in general and instruction in particular should be.

In Algeria, evidence from assessment practices in the school system is scanty and such evidence is not of quality to support the inferences that tests are expected to yield in relation to teaching and learning. It can be argued that the current testing system used by teachers in the Algerian educational system may be developed under pressure to reflect more closely, and to sustain desired educational goals. There is clear evidence that these practices are meant to exploit the format and content of developed tests to improve the final outcomes quickly and efficiently. In reality, there is something wrong in the way these tests are conceived. The assessment practices do not consider the fundamental theoretical and procedural constructs of tests and the connection that should exist between teaching and learning and the mechanisms by which students have to be assessed. In this spirit, it is assumed that the employed traditional testing system leads to inappropriate or outmoded forms of inferences which often fail to keep pace with the requirements of pedagogy.

Therefore, if improvement in the educational system is to be realised, the testing system should be used principally in support of teaching and learning, rather than in producing limited outcomes for the sake of just passing or failing. The underlying objective in this context should be employed in supporting of the learning process. Thus, the aim of assessment is said to develop and improve students' achievement and progress in learning the target language, and not just to measure their performances in some given skills and language components. Of course, this could be attained only when testing is connected to teaching and learning process through processes of feedback. In addition, there is a need that teachers should ensure appropriate standards of test design and development and content strategies with the intention to relate their tests more closely to valued classroom behaviours as these

are the underlying assumptions of the educational objectives set out in the official syllabuses suggested for the teaching of English language in the Algerian schools. In short, achievement tests should mirror the best practices of teachers, so that test practice will involve students in activities which will develop a full range of skills in the target language.

1. Statement of the Problem

From our experience of 10 years teaching English language at the secondary school level, we observed that the testing practices through the available assessment instruments did not provide specific available information about students' achievement and progress in English. The current testing system employed by the teachers in the Algerian schools relied only on either copies of the "Bac" exam model, or merely intuitive tests constructed without reference to any theoretical bases or operational procedures. Wrongly, what teachers attempt to do is to build test contents on the basis of previous existing tests, believing that such a practice would improve students' scores. It is ostensible that there is not much congruence between test-content and the contents of the syllabuses. The syllabuses, very often, are based on the integration of the four skills and the development of competencies; but, when it comes to assessment, our observation with regard to this issue has shown that tests assess one skill and/or two skills and eschew the other skills. For instance, the didactic unit presented to students throughout the learning cycle starts with a listening phase, but students are never tested on that skill. The same remark can be made about the speaking skill.

This study was stimulated by another observation in the field of language testing in the Algerian context. A dangerous phenomenon has cropped up: the scores have lost their credibility in judging the actual degree of achievement and progress of students. Those who might seem bright are really not the best while those with low scores are often believed to be weak, as far as the learning of English language is concerned. The interpretation of this phenomenon might reside in the misconception of language test design and development. As

a result, a negative washback effect has emerged. Instead of designing and developing a test that should assess the amount of mastery of the content of the syllabus, and to see to what extent the educational goals have been reached, teachers turned to becoming mere trainers of students on how to respond mechanically to a typology of questions and activities that are frequently included in the "Bac" exam papers. This system urged these teachers to considerably reduce the time available for instruction, restrict the range of the curriculum and limit the teaching methods, and potentially diminish their freedom to teach content or use methods that are believed to be incompatible with instruction.

As pointed out earlier in this section, having been an English teacher, what bothered us most about the English language teaching in the Algerian secondary schools is that teachers in this context do not realise how important assessment has been in shaping the current teaching and learning situation. We consider the situation one of a serious case of negative washback of an external standardized examination on the processes of teaching and learning.

It is within this framework that we conceived our work in order to understand the reasons of the set of anomalies observed in the current testing system in the context under exploration. The intention is not to criticize this system for the sake of criticizing it, but it is believed that this study would be an opportunity to judge the mechanisms by which the Algerian teachers come to evaluate and assess English language learners, and highlight common errors that currently occur in their assessment practices. To remedy this situation, we found it more appropriate to suggest an Alternative Testing Model (ATM) that seeks to overcome the deficiencies diagnosed in the current testing model, and to repair the shortcomings that are thought to be one of the sources of the decline in learning English language. To be able to succeed in this new project, there is a need to conceive the relationship that ought to exist between teaching, learning, and testing as one of partnership.

Each element in this equation has a tremendous role to play for the ultimate goals traced in this research.

2.Research Questions

Based on what has been stated in the problematic, the study explores these research questions:

RQ$_1$: What is the nature and scope of the washback effect on teachers' behaviours of aspects of teaching English language in the Algerian secondary schools due to the use of the current testing system?

RQ$_2$: What is the nature and scope of the washback effect on students' perceptions of aspects of teaching English language in the Algerian secondary schools due to the use of the current testing system?

RQ$_3$: What strategies does one need to implement the Alternative Testing Model with EFL classes at the Algerian secondary school level?

RQ$_4$: What is the nature and scope of the washback effect on teachers' behaviours as a result of the Alternative Testing Model implementation?

RQ$_5$: What is the nature and scope of the washback effect on students' perceptions as a result of the Alternative Testing Model implementation?

3. Aims of the Study

The general purpose of the present study is to explore how those involved in teaching and learning of English in the Algerian secondary schools perceive themselves to be affected by the implementation of the Alternative Testing Model. More specifically, this study aims at:

1. investigating the phenomenon of washback effect in light of using the ATM.

2. understanding how the main participants at the micro-level (teachers and students) reacted to the implementation of the ATM. In other words, the study offers the teachers and students an opportunity to voice their opinions to decision makers for a serious consideration of the issue of testing. Such a

purpose can provide a new solid ground for more consistent assessment reforms in the Algerian educational system.

3. exploring the nature of washback effect on aspects of teachers' and students' perceptions of the new model, and teachers' behaviours within the context of English in the Algerian secondary schools.

4. Hypotheses

On the basis of the research questions formulated before, we hypothesised that:

a. A useful achievement test for EFL secondary classes in Algeria will influence positively teaching and learning; and, conversely, an achievement test for EFL learners that is not useful will influence negatively teaching and learning.

b. A useful achievement test will influence attitudes towards the content and method of teaching and learning.

c. An achievement test that has important consequences will have positive washback; and, conversely, an achievement test that does not have important consequences will have no washback.

5. Research Methodology

Because the intention of this study is an exploration of the washback effect of an ATM on teaching and learning for EFL learners at different levels, focusing on perceptions, values, and situational factors in the complex and varying situation of the Algerian secondary schools, the design of the study needs to take into account all variables that are concerned with the different facets of teaching and learning where such a phenomenon may occur.

In order to draw a picture of the context where the current testing practices take place and find out the differences of this situation with the intended changes that might occur with the ATM implementation, the participants' perceptions, attitudes, and opinions were

significantly quested between two phases: (1) prior and (2) after the introduction of the new testing policy. Accordingly, a mixed-methods approach to data collection seemed appropriate; no single method is able to explain such a complex phenomenon as different methods have different strengths and weaknesses. Thus, using a range of methods can be considered the best way to understand the problem.

A mixed-methods approach which is both qualitative and quantitative data driven would allow the refinement and checking in context of these methods. As a result, the research methodology in the present study relies on questionnaires, interviews, classroom observations, and focus groups as data collection methods. These methods are seen to complement one another and could be relatively integrated in practice.

In order to verify the results and check for the validity of these results, the commonly known statistical method SPSS (Statistical Package for Social Sciences) is employed as an additional validating instrument in this study. From a research point of view, SPSS is useable as a complement to the used methods to corroborate the results. From a statistical point of view, the application of this method allows the researcher to validate the findings in some methods by cross-checking them with other methods.

6.Structure of the Thesis

Besides a general introduction and a general conclusion, the thesis is divided into 6 chapters. **Chapter 1** provides background information and the literature review on the central research concept of 'washback' in this study. It also displays its implications to innovation in education and provides a washback model for the present study.

Chapter 2 identifies the research methodology adopted in the current study. The chapter describes the research methodology, research strategies, and data collection methods.

Chapter 3 describes how the teacher-participants in this study perceive the current testing system they use to assess and evaluate their students' achievement and progress. In

this chapter, the research design highlights the research questions and purposes and delineates the procedures used to develop the research instruments in the Preliminary Study. This includes procedures for validating and increasing reliability of the data collection methods, and the procedures for collecting data.

Chapter 4 describes the proposed ATM, i.e., it displays the components of the new testing system. The content and format of the three sub-tests that form an achievement test based on the new model are discussed. The chapter ends with a description of the time framework and scoring procedure this model utilises.

Chapter 5 discusses the results obtained after the implementation of the ATM. The reactions of the participants to this innovation are shown in this chapter. As with the Preliminary Study, the design of the Final Study is displayed. This concerns the data collection methods used, the kind of respondents in each one of the research instrument, the data collection and analysis procedures. The chapter ends with discussion and summary of the findings.

Chapter 6 synthesizes the findings from the pre-implementation to the post-implementation time of the new testing model. In addition, the chapter devotes a section to the implications and recommendations of the study for those parties involved in the teaching and learning of English, and limitations of the study.

Chapter One: Washback in Language Testing

Introduction

The gist of this chapter turns around the underpinnings that shaped and guided this research. First, the chapter highlights the origins of washback as a recent concept that has come to emerge in the scenes of language education in general and language testing in particular. Second, it provides definition of this concept and its related constructs as it is conceived by most language testing specialists. Then, it identifies the different types of washback. Following this section, the chapter displays the functions and mechanisms of the concept under exploration. Next, washback as a lever to innovation and change in educational settings is discussed. Finally, this chapter ends with a potential washback model for the present study that investigates how the different components that make-up washback are structured. It also provides an ideal opportunity to understand how new testing systems are introduced into different educational systems.

1.1 Washback: Origins and Definitions

In recent years, washback has led to a greater understanding of this construct in the testing literature. In this section, a review of the literature related to the origins and definitions of this concept will be displayed. Besides, from this literature review, the different points of review about what the construct of this concept may encompass will be thoroughly discussed.

1.1.1 Washback: Origins

Although the subject of the effects of examination has long been discussed in the literature of General Education (Kellaghan et al., 1982; Vernon, 1956), and has been looked at from different points of view (Madaus, 1988; Fredericksen, 1984), it has been common in the testing literature that the concept of 'washback', as it is known now, has come to attract

the attention of testing an assessment researchers only at the beginning of the 1990's. Before, applied linguists used different terms to refer to the idea of examination influence. Some of these terms included, 'test impact' (Bachman & Palmer, 1996; Baker, 1991), 'systemic validity' (Messick, 1989), 'measurement-driven instruction' (Popham, 1988), 'curriculum alignment' (Shepard, 1990), 'backwash' (Biggs, 1993), and possibly other terms.

Among this set of terms, two of them dominated the scene of the issue of examination influence. Reference here is to the two concepts: 'backwash' and 'washback'. Testing specialists have quickly admitted that in order to avoid any sort of confusion in terms of the adoption and use of the appropriate terminology, it would be better to assume that these two concepts can be used interchangeably, and, therefore there is no need to place a clear distinction between, on the one hand, 'washback', and, on the other, 'backwash' in language testing uses and practices. In this sense, Hamp-Lyons (1997) corroborated the idea and noted that "washback is one of the set of forms that has been used in language education and language testing to refer to a number of beliefs about the relationship between testing and learning" (295). She goes on to add that "another set of terms is 'backwash', but it would appear that the terms 'backwash' and 'washback' are used interchangeably in the field" (ibid). To confirm this interpretation of the two concepts, Hughes (1993) points out that there is an interchangeable use of the two terms. He makes it more explicit when he states that "where washback comes from, I don't know. That I know is that you can find backwash in dictionaries, but not washback" (57). However, in another context, Cheng and Curtis (2005) prefer the term 'washback' and not 'backwash' since they think that 'washback' is the concept that is frequently found in applied linguistics in general and language testing literature in particular. In short, this brief reviewing of the distinction between 'washback' and 'backwash' leads one to believe that these two terms do not present any kind of difference, rather, the great majority of testing researchers who have dealt with this matter

have agreed upon the idea that both of the two terms stand for the same sense, and hence each one of them can be used interchangeably with the other one.

Language testing researchers have realized that the emergence of this concept is the result of the considerable reforms and advances that have taken part in the field of language testing during the last two decades at the end of the twentieth century. Indeed, it has been assumed that one of the areas that was actively discussed in that period of time was the influence of tests on both instruction and learning. Cheng (2005) indicates that the subject of examination influence was rooted in the notion that tests are often seen to drive teaching and learning. He argues that in order to realize what he refers to as measurement driven instruction, there is a dire need to seek for the creation of the matching between the construct of the test and what teachers present in instruction. In other words, he aims to say that the clearer the fit between test content and teaching is, the greater the potential improvement will be on the test. In a different approach, Messick (1989) placed the concept washback in a broader scope of construct validity. He claims that this construct encompasses a set of aspects about testing such as the impact of tests on language test takers and teachers, the interpretation of scores by decision makers and the intended uses of these scores. In such a view, the concept washback stands as an inherent quality of any kind of assessment, especially when test takers' futures are affected by examination results.

In a comprehensive study on how the concept washback has come to exist as an important research concept in the field of language teaching and testing. Tsagari (2006) proposed an artificial time framework divided into three different but successive phases: (i) the 'pre-1993' phase, (ii) the '1993' phase, and, finally, (iii) the 'post-1993' phase. First, he labelled the 'pre-1993' phase the 'myth' phase. He identified it as the period of time when writers recognized the examination influence phenomenon but no one accounted for it. What is noticeable in this era was that few empirical studies were carried out and published to the

language testing community (Wesdrop, 1982; Hughes, 1988), which made strong claims of the absence of this phenomenon. Most of the available studies in this period were merely based on self-report data or on direct results or on test results rather that direct contact with participants involved. Second, the '1993' phase, was markedly different from the previous one since it was typically characterised by the publication of a seminal work paper by two prominent language testing researchers, Alderson and Wall, who are indebted the fact they were the first who questioned the nature of examination influence. More crucially, the authors managed to reconceptualise this phenomenon by suggesting a set of relevant hypotheses. Finally, the third 'post 1993' phase, or as Tsagari named it the 'reality phase', was fundamentally recognized to be a new era where substantial research projects on washback used developed models to accurately decorticate and explain the various components that make up this concept.

To summarize, although relatively little has been written about the origins of washback, a great deal of information has emerged about various concepts that refer to examination influence. Based on this review, it is not an exaggeration to say that the study of the origins of washback is crucial to shape the scope of further needed research in this area. This matter should be treated as a direct consequence of other educational studies that targeted the investigation on the relationship between learning, teaching, and testing.

1.1.2 Washback: Definition

In testing research and literature, definitions of washback are numerous. These definitions vary from simple and straightforward to complex. Some take a narrow focus on teachers and students, while others transgress to concern even educational systems and society in general. Some definitions stress on intentionality whereas others insist they occur haphazardly (Bailey, 1999 : 3). In this subsection, a discussion of these definitions and how they are similar to or different from one another will be presented.

Many applied linguists have indicated that the concept washback is rarely found in language dictionaries. The few available definitions can be found in dictionaries such as 'The New Webster's Comprehensive Dictionary', which says washback as "the unwelcome repercussions of some social actions", expressed on the 'Collins Cobuild Dictionary of English', which defines washback as "the unpleasant after-effects of an event or situation". Apart from these two examples, a meticulous research in other common language dictionaries in English has shown that there does not exist an explanation or indication of the term washback as it is generally known in the testing literature in the present time.

Unlike the rare definitions found in language dictionaries, a great number of other definitions of the concept washback is present throughout the published assessment research and literature with various meanings. In a paper on testing listening comprehension, Buck (1988) describes the apparent effect of Japanese University entrance examinations on English language learning in Japan. In this description of washback, he says:

> There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of student, and pass rates are used as a measure of teacher success. This influence of the test on the classroom (referred to as washback by language testers) is, of course important; this washback effect can be either beneficial or harmful (17).

In this sense, Buck's definition emphasises the importance of what teachers and students do in classrooms.

In another definition, Messik (1996) considers washback as a prominent concept in applied linguistics. For him, this term refers to the extent to which the interaction and the use of a test influence language teachers and learners to do things they would not otherwise do that promote or inhibit language learning. Shohamy (1992) also focuses on washback in terms of language learners as test-takers when she describes, "the utilisation of external language tests to affect and derive foreign language learning in the school context" (15). She points out that "this phenomenon is the result of the strong authority of external testing and

14

the major impact it has on the lives of test takers" (ibid). To corroborate this belief, Shohamy cites the example of the introduction of an oral test proficiency based on an interview in the Unites States. She says that this example involves "the power of tests to change the behaviour of teachers and students" (Shohamy: 514).

Bailey (1999) refers to Shohamy who summarized four key definitions that are crucial to understand the washback concept. Accordingly,

1. Washback effect refers to the impact tests have on teaching and learning.

2. Measurement driven instruction refers to the motion that tests should drive learning.

3. Curriculum alignment focuses on the connection between testing and the teaching syllabus.

4. Systemic validity implies the integration of tests into the educational system and the need to demonstrate that the introduction of a new test can improve learning.

Cheng (2005) converged to some extent with Shohamy's (1992) ideas. She relies on Pearsons (1988) to show the influence of external examinations on the attitudes, behaviours, and motivations of classroom teachers, learners, and even on other broader, related areas of this research concept. Similarly, Cheng (1997) introduced the concept of intensity, "the degree of washback effects in area or a number of areas of teaching and learning affected by an examination" (43). Cohen (1994) also takes a broad view. He describes washback in terms of "how assessment instruments affect educational practices and beliefs" (41). In the same vein, Pierce (1992) suggests that "the washback effect, sometimes referred to as the systemic validity of a test, refers to the impact of a test on classroom pedagogy, curriculum development, and educational policy" (687).

However, a number of researchers in general education have reported that before yielding an accurate definition of the concept washback in its boarder sense, there is a crucial

need to clearly display the distinction between washback and the other confusing concept impact in language testing. On this matter, Tsagari (2006), in a study on washback, explores the relationship and/or distinction between these two major concepts: washback and impact. He argues that the common view which prevails in the field of language assessment considers washback as one dimension of impact. The latter is often used to describe effects on the wider educational context. Tsagari goes back to Wall (1995) to discuss in some detail the existing relationship between washback and impact. The latter suggests that "washback is frequently seen to refer to the effects that tests may have on teaching and learning, whereas impact deals with the effects that tests may have on individuals, policies, and practices, within the classroom, the school, the educational system, or even societies as a whole" (16). Following this interpretation, Tsagari (2006) recognizes that this view intersects to a great extent with other writers' explanations, such as McNamara (2000) and Shohamy et al. (1996) who place washback within the scope of impact.

These ideas are revisited in Bachman and Palmer (1996 : 35). They note that washback, however, is a more complex phenomenon than simply the effect of a test on teaching and learning. Instead, they feel that the impact of a test should be evaluated with reference to contextual variables of society's goals and values, the educational system in which the test is used, and the potential outcomes of its use. They referred to these uses at two levels:

- a micro-level, in terms of individuals who are affected by the particular tests uses, especially test-takers and teachers, and;

- a macro-level, in terms of society and educational systems.

Bailey (1996) adopts "a holistic view on washback, but prefer to consider overall impact in terms of 'washback of learners', and 'washback to the program', counsellors, etc." (263-264). For Cain (2005), although the two forms of washback and impact are used in many cases, interchangeably, the test impact more accurately refers to the wider implications

and effects of a given test use. Andrews (1994 : 37) writing on washback, appears to acknowledge the fragility of the washback-impact distinction. He reports that:

> The term washback is interpreted broadly ... washback refers to the effects of tests on teaching and learning, the educational system, and the various stockholders in the educational process, whereas the word process is used in a non-technical sense, as a synonym of effect.

Hawkey (2006) comments on the fact the concepts washback and impact are often considered in terms of their:

- logical location,

- definition scope,

- positive and negative implications,

- intentionality,

- complexity,

- direction,

- intensity, emphasis,

- relationship with validity and validation,

- relationship with testing critical view, and,

- the role in washback/ impact models.

Alderson and Wall (1993) also discussed the notion of washback, and tried to identify what washback was. The authors reviewed the concept as it has been presented by language specialists up to that time. They concluded that the concept was too vaguely defined to be useful and much of what have been said and written about this concept had been based on assertion rather than empirical findings. As response to this claim, they presented a number of 'washback hypotheses', which were meant to illustrate some of the effects that tests might have on teaching and learning. They argued that test developers should specify the types of impact that they wished to promote and the kinds of effects test evaluators should look for when deciding whether or not the desired washback has occurred (Wall, 2005: 51). The

washback hypotheses they presented stated:

1. A test will influence teaching.

2. A test will influence learning.

3. A test will influence what teachers teach; and

4. A test will influence how teachers teach; and

5. A test will influence what learners learn.

6. A test will influence how learners learn.

7. A test will influence the rate and sequence of teaching; and

8. A test will influence the rate and sequence of learning.

9. A test will influence the degree and depth of teaching; and

10. A test will influence the degree and depth of learning.

11. A test will influence attitudes to the content, method, etc. of teaching and learning

12. Tests that have important consequences will have washback, and conversely.

13. Tests that do not have important consequences will have no washback.

14. Tests will have washback on all learners and teachers.

15. Tests will have washback effects for some learners and some teachers, but not for others.

Alderson and Wall proposed these hypotheses as a result of their own extensive work in Sri Lanka and reviewing case studies conducted in Nepal (Khaniya, 1990), Turkey (Hughes, 1988), and the Netherlands (Wesdrop, 1982).

After reviewing numerous definitions of the concept washback, it is evident that this concept is open to a variety of explanations and that there are number of variables one needs to consider when conducting research on this subject. Crucially, what comes out from this discussion is that washback can be defined according to two major scopes. One following a

narrower definition which focuses on the effects that a test has on teaching and learning, and the other following a wider and a more holistic view of washback that transgresses the classroom to take into account the educational system and society at large, which can as noted earlier in this section would be more accurately referred to as test impact. In this connection, Hamp-Lyons (1997) summarises the situation and the terminology well. She finds that Alderson and Wall's limitation of the term washback to influence on teaching, teachers, and learning seems now to be generally accepted, and the discussion of the wider influences of tests is considered under the term impact, with the term used in wider educational measurement literature. In a similar view, the adoption of Bachman's and Palmer's definition, which refers to issues of test use and social impact as 'macro' issues of impact, while washback takes place at the 'micro' level of participants, particularly teachers and learners, sounds the most acceptable.

It is not within the scope of the present study to look in details at the wider implications of testing. Rather, in the context of this research, and since this work is exploratory in nature, the washback adopted will be primarily concerned with the area identified by Alderson & Wall (1993) and Bachman & Palmer (1997), i.e., 'washback to teaching' and 'washback to learning'. In other words, the researcher will adopt the narrow definition of washback focusing more on the washback at the micro-level that investigates the effects of a suggested ATM on teachers and students. Besides, he will try to be consistent in the use of the two terms 'washback' and 'impact'. In this respect, he uses 'washback' to cover influences of language tests on language learners and teachers, language learning and teaching processes and outcomes. In the same vein, he uses 'impact' to cover influences of language tests on stakeholders beyond language learners and teachers.

1.2 Types of Washback

Assessment studies have indicated that washback very often implies movement in a particular direction. This movement in a particular direction is an inherent part on the use of this phenomenon to describe teaching-testing relationship. Hsu (2009) referred to Pearson (1988) who pointed out that:

> Public examination's influence the attitudes, behaviours, and motivation of teachers, learners, and parents, and because examination often come at the end of the course, this influence is seen working in a backward direction, hence, the term washback (98).

Besides, washback has also been perceived as bipolar- either negative (harmful) or positive (beneficial). Messick (1996) cites Alderson and Wall's (1993 : 17) definition of washback as the "extent to which a test influences language teachers and learners to do things they would not necessarily otherwise do that promote or inhibit language learning" (Messik: 241). They add that "tests can be powerful determiners, both positively and negatively, of what happens in classrooms" (cited in Hsu, 2009 : 46-47).

Following this line of argument in this sub-section, regarding the two types of washback, the tremendous impact and power of testing on teaching and learning in schools and whether this washback exerts a positive or negative influence will be discussed in some detail.

1.2.1 Negative Washback

Negative washback is seen by testing researchers as the negative influence of tests on teaching and learning. Alderson and Wall (1993) point out that:

> A negative washback is defined as the undesirable effects on teaching and learning of a particular test. The tests may fail to reflect the learning principles and/or the course objectives to which they are supposedly related (5).

In this case, such tests will lead to the narrowing of content in the curriculum, instead of covering a definite content from what has been learnt in class. For Vernon (1956),

"teachers tend to ignore subjects in activities that are not directly related to passing examination, and testing accordingly alter the curriculum in a negative way" (18). Once again, it is logical that those tests may fail to create correspondence between the learning principles and/or the course objectives to which they should be related (Cheng, 2005 : 8). More dangerous, negative washback can substantially reduce the time available for instruction, narrow curriculum offering, and modes of instruction, and potentially reduce the capacities for teachers to teach content and use methods and materials that are incompatible with useful testing instruments (Smith, 1991 : 120). Madaus (1988) intersects with the above claims and adds that

> The negative washback definitely result in cramming, narrowing the curriculum, focus attention of those skills that are the most relevant to testing, placement of constrains on teachers' and learners' creativity and spontaneity, and disparage the professional judgment of educators (22).

One strong impression that resulted from negative washback is that an increasing number of coaching classes are set up to prepare students for examinations, but what students will learn are test- taking skills rather than language activities (Wiseman, 1961 : 21). In such a learning context, an atmosphere of high anxiety and fear of test results become current among teachers and learners (Shohamy et al., 1996 : 9). For Shohamy, teachers will feel that success or failure of their students is reflected on them, and they speak of pressure to cover the materials for the examination. When the students know that one single measure of performance can determine their levels, they will less likely take a positive attitude toward learning.

1.2.2 Positive Washback

There are other testing researchers, on the other hand, who have seen washback in a more positive way (Andrews, Fullilove and Wong, 2002; Bailey, 1996; Davies, 1985; Hsu, 2009). Those researchers strongly believe that it is possible to bring about beneficial changes

in teaching by changing examinations, representing the positive washback (Cheng & Watanabe, 2004 : 10). This phenomenon refers to tests and examinations that influence teaching and learning positively (Alderson & Wall, 1993 : 15). In a broad sense, good tests can be utilised and designed as beneficial teaching-learning activities so as to encouraging a positive teaching-learning process (Pearson, 1988 : 7). Andrews et al. (2002) suggest deliberately introducing innovations in the language curriculum through modifications in language testing. For instance, an oral proficiency test was introduced in the expectation that it would promote the teaching of speaking (Hsu, 2009 : 49). Davies (1985) considers that "a creative and innovative test can advantageously result in syllabus alternation or even in a new syllabus" (18). In this sense, a test no longer needs to be only an obedient servant; rather, it can also be a leader.

Nevertheless, in educational settings, things sound a little bit different as one may think of in that assessment researchers have come to realise that there exists a set of conflicting positions towards washback in language testing. That is, most of these experts claim that there is no clear consensus among practitioners as to whether certain washback effects are negative or positive. One justification to this conflicting situation is that potentially positive or negative nature of the test can be influenced by many contextual factors (Hsu, 2009 : 9). Alderson and Wall (1993 : 117-118), commenting on this particular case, posit that the quality of the washback effect might be said to have beneficial or detrimental washback. They add that whatever changes educators would like to bring about in teaching and learning by a particular assessment method, it is worthwhile to first explore the educational context in which an assessment is introduced.

Therefore, for many testing specialists research into washback may be more fruitful if this latter turns its attention looking at the complex causes of such phenomenon in teaching and learning, rather than focusing on deciding whether or not the effects can be classified as

positive or negative. According to Alderson and Wall (1993), the best way to realise this is to investigate as thoroughly as possible the broad educational context in which the act of assessing is taking part, since the major variables that often affect this act exist within the education system, and that might prevent washback from appearing. Cheng and Watanabe (2004) summarise this situation, and note that "if the consequences of a particular test for which teaching and learning are to be evaluated, the educational context in which the test takes place needs to be fully understood" (31). This means that whether the washback effect is positive or negative, this will largely depend on where and how it exists and manifests itself within a particular educational context in order to understand the mechanism of washback.

1.3 Functions and Mechanisms of Washback

Traditionally, tests used to be given at the end of the teaching and learning processes to provide an accurate diagnosis of the effects of teaching and learning. Nevertheless, with the advances and changes made in the field of testing and how this latter is conceived, a test can also be developed to be used at the beginning or in the middle of the teaching and learning processes in order to influence either or both processes. This section intends to shed light on the functions and mechanisms by which washback occurs in relation to other educational theories and practices.

1.3.1 Functions of Washback

In discussing the functions of language tests through which washback occurs in actual teaching and learning environments, Wall (2005) referred to a set of reviews of those tests and influences they could have on the systems they are introduced into. One of these crucial reviews is the one that was produced by Eckstein and Noah (1993). In its essence, Eckstein and Noah provided a historical account of the myriads of a number of functions and influences of some types of tests that displayed appropriately how people over history have

usually considered tests as an important tool by which they take the desired decisions for some targeted purposes. For instance, for the authors, the first documented use of written, public examination systems occurred under the Han Dynasty in China about 200 B.C. The main function of these particular examinations was to select candidates for entry into the government services. In other words, the candidates were used to break the monopoly over government jobs enjoyed by an aristocratic feudal system.

With Eckstein and Noah, the second example of the functions of tests was that one which sought to check patronage and corruption. A typical example of this function was Britain where people could gain entry into higher education or the profession of strengths. An important, direct consequence of this examination was the establishment of a great deal of public schools, which aimed at preparing students to sit for these examinations. In addition to this, a third example of functions of examinations, suggested by Eckstein and Noah, was to encourage levels of competence and knowledge amongst those who were entering government services or professions. The intention was to design and develop examinations which reflected the demands and requirements of the target situations; students for those examinations could have to develop skills which were relevant to the work they hoped to get in the future.

The fourth function, in this series of examples, was that of allocating spare places in higher education. At this level, examinations were used as means of selecting the most able candidates for the available places. This type of examinations is the same to what is referred to as placement tests in the testing literature in the present time. The fifth function in this illustration was to measure and improve the effectiveness of teachers and schools. Eckstein and Noah used again Britain, as an example describing how, at a certain time, the government set up a system of examination to monitor performance of schools by sponsoring these examinations through the allocation of considerable funds. The amount of funds that the

school received depended on how its students performed. However, the system had serious unintended consequences and at last had forced to achieve the expected objectives. The last function, in this set of examples suggested by Eckstein and Noah, was limiting curriculum differentiation. In Britain, in the nineteenth and the twentieth centuries, there was a remarkable resistance to the idea of centralised education, and all the schools had the freedom to decide on their own curriculum and means of assessment. With the establishment of certificate examinations, these schools had a common target they could aim for, and all these schools turned to teach the curriculum that can help better in doing well in the examinations that are relevant in these certificates.

In the modern world, tests are frequently used for accountability within the system, and in particular for certification of achievement in education. They form part of the procedure for decisions about the allocation of scarce resources of both systemic and an individual level. For example, tests in many countries and Algeria is one of them, tests control the transition between school and higher education, and they may lead to the awarding of a degree. Tests are also seen as ways to upgrade knowledge and to improve the performance of institutions (Hsu, 2009 : 62). Through testing, education policy can be rapidly defused and implemented at relatively low cost (Linn, 2000). Test results that are visible and ideally measurable can be reported by the media in terms the public can understand and can be used to show that change has or has not taken place. However, tests are also criticized for exerting a certain authority and power on both systemic and individual level. But, in spite of the criticism levelled at them, tests continue to occupy a leading place in the educational system of many countries.

The series of functions, exposed above, are typical situations where these tests were used to exert influence on the final outcomes to suit the expected intentions of those who are in authority to make and impose their policies. As pointed out by some testing specialists, this

is an especially common practice in countries with centralized educational systems, where the taught programmes are controlled by central agencies. Policymakers in these contexts and countries have used tests to manipulate educational systems, to control curricula, and to impose new textbooks and teaching methods. In such settings, tests have been viewed as a primary tool through which changes in the educational system can be introduced without having to change other educational components such as teaching training and curricular. On this point, Shohamy (1993) commented on that:

> The power and authority of tests and external examinations enable policy-makers to use them as effective tools for controlling educational systems and prescribing the behaviours of those who are affected by their results - administrators, teachers, students and others (239).

Given the status of tests and examinations in public spheres, it seems that it is important to understand the functions of testing in relation to many facets and scopes of teaching as mentioned in the examples discussed earlier. The importance of considering these functions serves as a starting point and also a linking point to get a clear picture of the various mechanisms within different educational contexts.

1.3.2 Mechanisms of Washback

In explaining the complex mechanisms through which washback occurs in actual teaching and learning environments. Bailey (1996) cited Hughes (1993) trichotomy to show how this phenomenon works in different contexts. Bailey points out that this particular trichotomy allows educators in general education and language testing specialists in particular to develop a basic model of washback that explains how the various components that make-up this framework interact to help the understanding of the nature of this subject of interest. In describing this model, Hughes states that the trichotomy is formed of three parts. First, the participants who are mainly the people such as classroom-teachers, students, administrators, materials developers, and even publishers whose perceptions and attitudes

26

toward their work may be a test; Hughes' second component in this framework is termed process. The latter covers any actions taken by the participants, which may contribute to the process of learning as the development of teaching methods. Third, in Hughes' framework, product refers to what is learnt as facts, skills, and other aspects and also the quality of learning.

Contrary to Hughes who stresses more on the three component that make-up this model, Alderson and Wall (1993), in their Sri Lankan study, focus on what they referred to as 'micro-aspects' of teaching and learning that might be influenced by examinations. Alderson and Wall argued that there is little evidence provided by empirical research to sustain the idea that tests impact on teaching. They advocated that:

> The concept is not well defined, and we believe that it is important to be more precise about what washback might be before we can investigate its nature and whether it is a natural or inevitable consequence of testing (117).

Consequently, they suggest 15 hypotheses that can aid researchers to illustrate areas in teaching and learning that are usually affected by washback and can stand as a basis for further researcher. This set of hypotheses has shown that there exists a strong correlation between the importance of tests and the extent of washback. Alderson and Wall concluded that further research is needed and that such research must "entail increasing specification of the washback hypothesis". They called on that researchers in the field of language testing had to take into account to research literature in at least two areas: motivation, performance, and that of innovation in the educational settings.

Following this seminal work realised on washback hypotheses, Wall (1996) followed up their study and stressed the difficulties in finding explanations on how tests exerted influence on teaching. She went back to innovation theory and literature to explore the complex topic of washback. In this respect, she proposed that the research areas that are seen to be relevant to washback should include (a) the writing of detailed baseline studies to

27

identify important characteristics in the target system and the environment, including an analysis of the current testing practices (Shohamy, Donista-Scmidt and Ferman, 1996), current testing resources (Bailey, 1996; Hughes, 1993);

(b) attitudes stockholders (Bailey, 1996), and;

(c) formation of management terms representing all important interest groups; teachers, teachers-trainers, ministry officials, parents, and learners (Cheng, 2008).

Likewise, in the same perspective of washback mechanisms as a phenomenon of change in teaching and learning, Hsu (2009) referred to Smith (1991) who investigated an ELT project and worked on to construct a corresponding model of variables involved with the aim to introduce the desired change in the teaching and learning processes. In its essence, smith's model comprises five components of change: the target system, the management system, the innovation itself, the resources available, and the environment in which change is supposed to take place. Hsu adds that, on the ground of the same idea, Markee (1997) illustrated through another study how change might occur on larger subjects such as curricular through following stages which are to design, to implement, and finally to maintain. In this respect, Markee suggested a framework that was based on the composed questions that were posed by Cooper (1989) and which referred to: **who** (participants), **what** (product), **where** (the content), **when** (the time, duration), **why** (the rationale), and how (different approaches in managing the washback effect).

In two other studies, Fullan with Steingelbawer (1991) dealt with the issue of washback effect but in its broader uses. They discussed the effects and changes of tests on schools and came to identify two main recurring themes: first, a washback effect should be seen as a process rather that an event. Second, all participants who are affected by this phenomenon have to find their understanding of what washback effect is. Cheng (2004) made this last point clearer. He explained that according to Fullan teachers work on their own with little

reference to experts or consultation with colleagues. Thus, those latter are usually forced to make on -the-spot- decisions, with little time to reflect on their better solutions. The other problem they often encounter in this context is that they are always unable to accomplish what they prepared to do. Consequently, their lives can become very difficult, indeed. This reality can explain why intended washback does or does not occur in teaching and learning. In other words, this means that, if educational change is often imposed upon teachers and students without meticulous preparation, resistance is likely to be a natural response (Curtis, 2000 : 4).

In summary, these reviews of major studies on the mechanisms of washback have corroborated the fundamental relationship between the design of given tests and their positive or negative impact and power on teaching and learning. However, it is worth noting that the outcomes of these studies, even if they have contributed in advancing research into the domain of washback in language testing, they remain insufficient to draw a larger and transparent picture of this issue since a number of raised questions on the mechanisms of washback in language testing remain without definite answers.

1.4 Washback: Empirical Studies

In this section, a number of common empirical research studies into washback of both language and general education are discussed. This literature review is a summary of detailed reviews realized by a number of researchers. The latter highlighted the basis of the central research concept and pointed out useful research methods adopted by the myriads of researchers to carry out their investigations. Such an elucidation is of a great utility for the present exploration since it shapes the scope of the study and serves as a guideline for many relevant issues for further needed research. For ease of reference Table 1.1 provides background information for the most used studies in terms of the educational context, exam type, and research methods employed.

Authors	Context	Exam	Methods
Wesdrop (1982)	The Netherlands	Multiple choice language assessment and final exams in Dutch Secondary Schools.	• Scores • Analysis of tests • Teacher and students' questionnaires.
Hughes (1988)	Turkey	University entrance test.	• Test scores • Questionnaires to lecturers.
Li (1990)	China	Matriculation English test (MET)	• Questionnaire to teachers and local officers (and students' discussion).
Alderson and Wall (1993)	Sri Lanka	O- level examination in English	• Individual and group in interviews with teachers • Questionnaires to teachers and students and teacher advisors • Materials and test analysis • Observations
Lam (1993)	Hong Kong	New use of English (NUE) (end of secondary school)	• Questionnaire to teachers • Test book analysis • Analysis of test script and scores
Andrews (1994)	Hong Kong	Oral component of the revised use of English (RUE)	• Two parallel questionnaires to the working party members and teachers
Alderson and Hamp-Lyons (1996)	USA	TOEFL Exam	• Individual and group teacher and student interviews • Observations • Fieldnotes

Watanabe (1997)	Japan	University entrance exam	• Questionnaires • Interviews with students and teachers
Read and Hayes (2003)	New Zealand	IELTS	• Interviews • Questionnaires • Observations • Pre- and Post-English test
Qui (2004)	China	National Matriculation English Test (NMET)	• Interviews and questionnaires with NMET constructors, inspectors, teachers, and students • Observations
Cheng (2005)	Hong Kong	Revised Hong Kong Certificate of Education Examination (HKCEE)	• Questionnaires to teachers and students • Observations • Interviews
Gosa (2006)	Romania	English component of the Romanian school learning exam (Bac)	• Students diaries (10 students retrospective use)
Hawkey (2006)	UK	CPE (Cambridge EOSOL)	• Text book analysis

Table 1.1: Overview of the common research projects into washback

Referring back to the above illustrative table, what is ostensible is that most reports from the various, available research studies into washback indicate that the influence of washback has been observed on various aspects of learning and teaching, and that this phenomenon generated was mediated by numerous factors. What is more significant on this matter is that almost all research projects looking at washback have been carried out in several different countries and various contexts. Crucially, all these research studies are organised with regard to what Hughes (1993) referred to as process-washback on content, teaching methods, and classroom assessment, product-washback on students learning, and participants-washback on feelings and attitudes of teachers and students in the context under

exploration. Thus, in order to have a clear understanding of how these research studies were carried out, it is worth examining several of the most interesting tenets of these research projects.

Wesdrop (1982) investigated whether the incorporation of multiple-choice technique in secondary school institutions would lead to the impoverishment of the curriculum. He argued that such a fear could happen because the skills that could not be tested through multiple-choice would not be practised and hence would eventually completely disappear. He added that if this happens there would be a failure in the adoption of some teaching methods. Also, this fact may provoke some changes in the way in that students prepared themselves for tests. Wesdrop concluded that his study revealed that there are no differences between teaching practices and students' preparation methods. He concluded with the assumption that after all the so-called washback effects are a mere "myth". If they do exist, they must be so weak or just small that our methods cannot detect them.

Hughes (1988), in a project at Bogaziri University in Istanbul, Turkey, explored how the introduction of a test in English for academic purposes helps to improve English performance in the university's English medium undergraduate courses. Hughes demonstrated that the basic aim of this test was to devise a new proficient test as the sole test by which students could get access to undergraduate programmes. The intended test was developed after completing a needs analysis. It comprised sections on three major skills: listening, reading, and writing. Hughes, after introducing this test, noted that, "for the first time, the foreign languages schools teachers were compelled by the test to consider seriously just how to provide their students with training appropriately for the tasks which they could face them at the end of the course" (44). Hughes concluded that the washback effect occurred as a result of the incorporation of this test, with great changes in the materials used in the foreign language school.

Likewise, Li (1990), in another washback study examined high-stakes examination taken by Chinese students at the end of secondary school, the so-called the matriculation English test-MET-. He pointed out that this test was first introduced in 1984 to replace an earlier examination, which was so weak and lacked considerable validity and reliability. Li aimed at identifying whether or not the introduced test would lead to better results or not, in comparison to the previous adopted examination. After a period of time from the incorporation of this new test, Li came to the conclusion that the teachers' and students' attitudes towards the MET were positive. In this respect, he wrote "tests are able to subjugate the minds of millions of people to the thraldom of forced memorisation, but, we would say it is a greater kind of power to be able to liberate people's minds from such thraldom" (42).

From the short discussion of these three research projects into washback illustrated above, the common view is that the very few empirical studies that were carried out prior to the 1990's period indicated that little effect and impact of tests on learning and teaching was found on classroom assessment. Such a result would be explained by the fact that these studies into washback research failed to construct a definite washback model that would take into account the array of factors which may place a part in determining why teachers react in the way they do.

In brief, the washback studies that prevailed in this particular phase succeeded to provide a clear definition of the concept washback, some guidelines about how to achieve positive washback, and a few references to the effects of tests on the contents they had been introduced into; but, in the meantime, there were few detailed accounts of specific attempts to innovate through testing. Most of the research was based on questionnaires or on test results rather than direct classroom observations.

Contrary to the preceding period of time, starting precisely from 1993 the scenes of language education in general and testing in particular have known a significant increase in empirical studies on washback effects. A great deal of language testing researchers recognise that the recent research projects have led to a more detailed understanding of the phenomenon in the domain of language education and of the factors which contribute to it.

Wall and Alderson (1993), for example, examined the effects of the new O-level examination on English teaching and learning in secondary schools in Sri Lanka. They emphasized that, by the time the study was published, it was the only investigation that included classroom observation as one of its research methods. Wall (1996) summarised the findings of the study by pointing out:

> The examination had had considerable impact of English lessons on the way teachers designed their classroom tests, but, on the other hand, this examination had had little to no impact in the methodology they used in the classroom or on the way they marked their students' performance (348).

Wall and Alderson found that the potential factors which impeded teachers from using new teaching methods included insufficient teacher training, problematic management schools, and teachers' beliefs in the efficacy of various methods.

In another context, Lam (1993) examined the New Use of English (NUE) examination in Hong Kong. Lam through this study, attempted to find out clear answers to a set of some raised questions, such as whether the amount of time that schools allocated to the teaching of English language is sufficient, the schools set aside special time to prepare for one particular section of the examination, how the attitudes and abilities of their students are, how the quality of English textbooks is, and how the content of the teaching and the students' performance are. He concluded that it is worth noting for the examination designers to take into account how different factors in the context where this examination occurs might interact with one another to yield the appropriate intended results and a clear picture of the expected examination.

Andrews' (1994) Hong Kong study, was about the development of the RUE -the Revised Use of English- test to measure students' oral performance in Hong Kong. In order to see the degree of efficiency of this newly, developed test, Andrews conducted a study using two parallel questionnaires to the working party members and teachers with three groups of candidates. The results of this investigation were that there was not one definite conclusion to the washback effect of the designed oral tests; rather, Andrews remarked that the final outcomes indicated that the nature of washback varied across the three groups: only a small improvement in performance between the first and the second group was ostensible. These results led the researchers to conclude that the washback effect of the test was delayed. For this reason, the findings of Andrews' study suggested to re-use the test in a second year to see whether or not the expected results could be noticeable.

Alderson and Hamp-Lyons (1996) found, in their study of washback of the test of English as a foreign language (TOEFL) on preparation courses, that "this particular test was seen to have a more direct washback effect on teaching content than on teaching methodology". The researchers employed three different types of data: interviews with students in groups, interviews with teachers (both individuals and groups), and field notes and audio-recordings during classroom observations. Like Watanabe (1996), Alderson and Hamp-Lyons observed "two different teachers while they taught both TOEFL preparation classes and courses. This particular design permitted Alderson and Hamp-Lyons to compare TOEFL preparation to non-TOEFL preparation classes" (cited in Bailey 1999 : 32). The authors concluded that the amount and type of washback which occurred depended on

> The status of the test, the extent to which the test is counter, the current practice, the extent to which teachers and textbook writers think about appropriate methods for test preparation, and the extent to which teachers and textbook writers are willing to and able to innovate (296).

A similar research design was used by Watanabe (1997), who investigated the university entrance examination in Japan through two different types of data collection methods: questionnaires, and interviews with teachers and students. Watanabe found that all the textbooks used by the teachers observed consisted of past exam papers and materials. In addition, the results showed the presence of grammar translation questions on a particular university entrance examination did not influence the teachers in the same way in that some teachers were affected by these exams, and others were not. Watanabe identified that "three possible factor that might promote or inhibit washback to teachers: (1) the teachers' educational background and/or experience; (2) differences in teachers' beliefs about effective teaching methods; and (3) the teaming of the researcher's observation" (cited in Bailey 1999 : 23) . Thus, Watanabe concluded that "teacher factors may out weight the influence of an examination in terms of how an exam preparation courses are actually taught" (ibid). Moreover, he noted that school cultures might influence the degree of washback in that "a school positive atmosphere which encouraged students to interact with authentic language might infiltrate into individual classrooms" (Hsu 2009 : 58).

Cheng (2005), in a large-scale empirical quantitative and qualitative study, sought to corroborate the idea of whether the modified Hong Kong Certificate of Education Examination (HKCEE) taken by most secondary graduates brought about the positive washback on teaching that was intended. In this study, Cheng used questionnaires, interviews, and observations during the first year after the announcement and discovered that the newly introduced examination was having considerable influence to the 'what' teachers teach, and not to the 'how' they teach (Wall, 2005). In other words, the charge of the examination would change teachers' classroom activities, but it did not change teachers' beliefs and attitudes about teaching. Cheng suggested that, "to change the how ... genuine changes in how teachers teach and textbooks are designed must be involved. A change in the

examination syllabus itself will not alone fulfil the intended goals"

In a study that converged with the findings found in Cheng's work, Read and Hayes (2003) attempted to measure the learning outcomes through an investigation conducted on the IELTS in New Zealand. The researchers used four data collection tools: interviews, questionnaires, observations, and pre- and post-English test. What was particular in this study was that the researchers had two small groups of 17 students. Those latter took retired version of the IELTS exam as a pre- and post-tests to two ILTS courses (intensive and general). Like the final result obtained in Cheng, Read and Hayes study did not show any significant improvement overall, nor between the groups of students. The researchers concluded that time is needed for washback to occur.

Furthermore, in order to determine how students' individual differences can be affected by washback effect, Ferman (2004) examined the influence of the introduction of an oral test on learners' achievement. The author found that average ability level students were significantly different from other students: their anxiety level was the lightest and they were not adversely affected by potential failure in the test. For that reason, Ferman concluded that in order that washback could occur, it is important to consider the individual differences among learners.

Like Ferman (2004), Gosa (2006) sought to identify possible washback effects that took place inside and outside classrooms as experienced by her students in Romania. Gosa used students' diaries to analyse whether or not the students' study environment was affected by test washback. Adopting that particular method, Gosa recognised that the individual differences among the learners and the environment where they operate need to be considered to see if an exam might exert the expected effects, attitudes; perceptions, beliefs, learning styles, and anxiety should be taken into account when trying to promote positive washback as they are likely to interact with the test, and hence intervene in the washback process.

Qui (2004) conducted a survey to examine the impact of the National Matriculation English Test (NMET) in China. This investigation focused on the main function used to select students for higher education. The obtained data revealed that the NMET has a considerable impact on materials and learning activities, but not the type of the intended results set out at the beginning of this survey. Qui concluded that one of the reasons for this was that teachers failed to teach students the required skills that are supposed to be an integral part of instructional objectives. Instead, these teachers felt more pressured to work only for good results that have to be obtained at the end of this examination.

Wall and Horâk (2006) examined the impact of the changes of the TOEFL on teaching and learning in preparing students to take the test from teachers' point of view in central eastern Europe. Wall and Horâk used interviews over the period of five months to detect the degree of answers of teachers on the changes in the TOEFL test. They observed that there was indeed a certain awareness, but it grew very slowly. Nevertheless, the two researchers found that there was a positive impact toward the introduction of a speaking test and the integrated writing skill. They concluded that the availability and quality of the information about the test and test preparation materials would be a major source contributing to teachers' reactions to desired changes.

Therefore, from the above literature review on washback effects a number of findings have emerged with regard to this phenomenon and the ways in which it can be investigated. Some of these concluding remarks are summarised as follows:

First of all, the review of the literature showed clearly that washback is broad and multi-faceted and can be brought through the agency of many independent and intervening variables besides the exam itself. As far as washback is concerned, one can see now some of the factors which seem to have affected the form that washback can take included teachers and students factors such as beliefs, attitudes, experience, education, training, personality, the

status of the subject to be tested, resources, classroom conditions, management of practices in the schools, communication between test providers and test users, and even the socio-political context which the test is put to use. In addition, what stands out clearly is that to carry on washback investigation, a multitude of different data collection methods were employed. Example of these used instruments concerned methods as classroom observation, individual interviews, group discussion, questionnaires, analysis of participants' diaries and their talk in the context under exploration. All these methods and instruments aimed to examine the factors and variables that make-up washback.

As was seen in this review of literature, the majority of research studies into washback tried to report the effects of examinations on the teaching content (Alderson & Hamp-Lyons, 1996; Read & Hayes, 2003; Wall & Alderson, 1993). Some results have indicated that tests altered teaching methods and materials, but others have shown that the tests had limited or no impact on either (Alderson & Hamp-Lyons, 1996; Cheng, 1997; Wall & Alderson, 1993). Washback may also be differential, it occurs with some teachers, but not with others (Alderson & Hamp-Lyons, 1996; Watanabe, 1996). Besides, most of the analysed data revealed that tests have a superficial impact on students learning, and these individual learners like teachers have experienced this influence in different ways, with the potential for considerable impact in terms of effective factors and teachers' behaviours (Cheng, 1997; Ferman, 2004). The differences between the degree of washback on teachers and students have raised questions about the extent to which washback to teachers can be assumed to be generalizable to washback to learners. In sum, there is an evidence indicating how washback to teachers and programmes might interact with washback to learners.

At the end, on the ground of the literature review dealt with above, there has been an evidence displaying most of the common and available empirical studies that investigated washback in language testing on learning and teaching but at large-scale proficiency tests. In

other words, this implies that a study of washback on assessment in a classroom context is still not well explored. This reality is expressed by McNamara (2000), too much language testing research is about high-stakes proficiency test, ignoring classroom context, and focusing on the use of technically sophisticated qualitative methods to improve the quality of tests at the expense of methods more available to non-expert. Hence, what is required to overcome this shortage is to realize a study of washback on learning and teaching for classroom tests, and the influences of these tests on the teachers' and students' behaviours and attitudes. This objective becomes the priority of any washback investigation, and this is actually the main argument on which the present study in this thesis rests.

1.5 Washback: Lever of Innovation in Education

Washback is the key concept in this study. So, this section will attempt to clarify the implication of this concept to educational innovation. The aim of this section is to highlight a possible framework that can help test developers to judge whether their innovations (the testing they are developing) are likely to have the impact they intend them to have. In the literature, it was argued that to understand the nature of washback it is also important to take account of findings in the research literature in the area of innovation in language and change in educational innovation.

A great deal of applied linguists assert that there has been a well-established tradition, which led to the realisation of a number of networks that served to yield the most elegant compilation of ideas about the different phases in the innovation process at the factors at work in every phase (Rogers, 2003; Fullan, 2007), and an increasing body of literature focusing on the English teaching context (Henrichsen, 1989; Kennedy, 1990; Markee, 1993, White, 1993; Li,2001). Crucially, what is most remarkable with these research studies is that they succeeded to some extent to make clear for readers the complexity of the innovation process, and the factors which inhibit or facilitate successful diffusion and implementation.

Because it is not possible to cover all of innovation theory in a single action, the discussion will be limited to the ideas which are relevant to the present study. The particularity of what is going to be displayed in this section is that ideas are arranged in a certain way in order to help readers find a link between innovation and washback in education. The researcher looks first at the term innovation and what it implies as a specific concept in relation to washback in language testing, then he considers what distinguishes this term from the other types of change needs to be considered. This will be followed by a discussion of the process of innovation and the sense of change, for the individuals who are most affected by it. This section concludes with the provision of several models of innovation, including the hybrid model of the diffusion/implementation process by Henrichsen (1989), which served as the starting point for the analysis of data in the present study.

1.5.1 Innovation: Definition

The first question that needs to be answered is what the term 'innovation' refers to. Following wall (2005: 60), Rogers (1985) defines innovation as an "idea, practice, or object that is perceived as new by an individual or other unit of adoption". For Rogers, he sees that:

> It does not matter whether the idea is objectively new (in terms of the amount of time that has passed since its discovery or invention), but rather whether it is felt to be new by those who may be adopting or using it (ibid).

In Hsu (2009), innovation can be usefully defined as a planned and deliberate effort, perceived as new by individuals to bring about improvement in relation to desired objectives. Hsu makes this idea more explicit. He advocates that educational innovation is the result of a number of problems that a given educational system can present such as failure in students' achievement, a poor performance by students in specific areas, or a lack of transparent accountability reporting. What is worth noting about these problems is that they also transgress to touch some aspects of educational systems that concern systematic attempts by

some authorities to change educational policies and practices with the intention to achieve better final outcomes (Brindeley, 2008 : 36).

Some other researchers make a distinction between innovation and other types of change. For Wall (2005: 60), who cites White (1993), "the difference has to do with intentionality: while 'change' is any difference that occurs between time one and time two, an 'innovation' requires human intervention" (244). For Miles (1964) "innovation is a deliberate, novel, specific change, which is thought to be more efficacious in accomplishing the goals of a system" (cited by White, 1988: 211). Nevertheless, there is another view borne by some other researchers who use the terms as synonyms. Many of them argue that if they believe that the distinction is a valid one, they use them interchangeably to avoid any sort of confusion or ambiguity, the researcher prefers to opt for the view that regards that these two terms bear the same sense since the efforts required to launch innovation in this research are so high, and the challenges go beyond the discussion between the two concepts.

In the literature on innovation in language education, it is thought of that the ideas provided by Markee (1987) are said to be the most comprehensive. Markee recommends that language-teaching professionals should adopt a 'diffusion -of-innovation perspective' in order to understand why their attempts to innovate meet with success or failure. In other words, this means that specialists need to be aware of the matters and findings reported by educators and vice versa. For Markee, this approach will not only provide language educators with a coherent set of guidelines and principles for the development of their own innovation evaluation, but will also apply them with criteria for retrospective evaluation of the extent to which these innovations have actually been implemented (cited in Wall, 2005 : 61).

What comes out from this brief discussion on the right sense of the concept 'innovation' is that it is deliberate, intentional and also planned. With regard to this definition, it is apparent that the other concept 'washback', the central research concept in

this study, should be conceived with a meaning that overlaps to a large extent with this definition of innovation in order to bring about changes and improvements to the teaching and learning processes. Obviously, this is what it is intended by the innovation of a newly, testing system in the context under study in this thesis. In what follows, more understanding about the process of innovation and what it encompasses, and how it is adapted to this research will be discussed.

1.5.2 The Process of Innovation

The concept of innovation has been so far defined. The next step in this section is to synthesise the process of innovation through the discussion of four major innovation views provided by four different educationists in order to highlight four attempts of innovation to bring about their relevant ideas together: a basic model for innovation by Rogers (1995), a comprehensive survey of innovation by Fullan (1991), a set of principles that are seen to guide an innovative act by Markee (1997), and Henrichsen's (1989) innovation model. The latter is discussed in some detail because of its relevance to language education and to language testing projects in various contexts.

1.5.2.1 Rogers' View

In the innovation process, a number of attributes have been proposed as correlating with the success of implementing innovation. One of the most cited set of attributes is perhaps that proposed by Rogers (2003). For Rogers, there are five attributes that compose his model, which are relative advantage, compatibility, observability, trialibity and complexity. The first attribute, as Rogers posits, is about the answer to the question that is first raised on the persons whom are mostly affected by innovation. In addition, relative advantage represents the perception that those persons have on the innovative act. It is believed that the greater an individual perceives the relative advantage to be, the quicker it

will be adopted. Hsu (2009) refers to Rogers to make this point more explicit:

> He linked relative advantage to a person's incentives. If incentives contribute strongly to a decision to adopt a change, there may be little relative advantage to its continued use after the incentive is removed or sometimes reduced (70).

Compatibility, the second attribute, provides a thorough description of the degree of congruence between the innovation and the existing values, past experiences and perceived needs of those who are expected to adopt this innovation. For Rogers, it is clear that if there is a high degree of compatibility between the innovation and the standard norms and values of a system where innovation is to occur, the act of innovation is going to happen rapidly. Contrary to this attribute, the third one, complexity, is proposed by Rogers to display the extent to which an innovation is difficult and complex to take. It is argued that if an innovation is seen by its adopters to be difficult, it will be difficult to diffuse and adopt (Rogers, 2003).

The next attribute in this series of Rogers' model concerns mainly the issue of triability. The latter refers to the extent to which a prospective adopter could try out an innovation before its adoption. For Rogers, an innovation that is triable represents less uncertainty to the person who is considering for its adoption, as it is possible to adopt the innovation a little at a time rather than all at once and to learn by doing. The final attribute, observability, pertains to the adopter's ability to actually see the innovation being used by others. For Rogers, if an innovation is observable, it is easy to adopt and defuse (Hsu, 2009 : 70). What is significant about this model is that it claims that considering these five attributes makes it easy for the innovation to be adopted and rapid for its diffusion Among these five attributes, innovation researchers believe that relative advantage and compatibility are of a great value, and very important in giving a thorough explanation of an innovation's rate of adoption

1.5.2.2 Fullan's View

On the issue of quality, Fullan and Steingelbawer (1991), in their comprehensive survey of innovation in education, observe that innovation should be regarded as a process rather than an event. For Fullan (2007), the innovation process is identified through three fundamental stages, which are: 'Initiation', 'Implementation', and 'Continuation'. What is worth noting on these stages is that Fullan recognised that it is not possible to predict how events in one phase will influence those in the others, or how long it will take for one phase to change into another.

Referring back to these three basic stages in Fullan's model, these are defined as follows:

1. The 'Initiation' stage: it is the process that occurs between the first appearance of an idea for a change and the time when it is adopted. In this stage, Fullan proposes to ask a set of questions to see whether or not the idea to be adopted is worthy. These questions include:

   - What is the source of the idea of change?

   - What was the motivation behind the idea?

   - What is the quality of the innovation?

   - Are the participants in the innovation process aware about the requirements of the idea?

   - Is this particular idea supported by external establishments?

2. The 'Implementation' stage: it is the process of putting into practice an idea, programme, or a set of activities and structures new to the people attempting or expecting change. Fullan insists on the assumption that there should be a definite consideration to the factors, which are important in this stage. This particularity includes three aspects: the characteristics of the innovation (need, clarity, and availability/practicability), the characteristics of the local context (the district, the community, the principal and the

teachers), and the characteristics of external bodies such as (government and ministries) (cited in Wall, 2005 : 62).

3. The 'Continuation' stage: it refers to whether an innovation becomes part of the educational system, or whether it fails and is rejected. Like the previous stage, a number of factors need to be considered. These mainly concern matters such as:

- the degree to which an innovation has been built into the system;

- the number of people who are committed to and skilled in the change;

- the strength and procedures to provide continued support, and

- the degree of the staff turnover in the target situation.

Besides the identification of the three stages of the implementation process, Fullan (2007) observed that quality in implementing projects needs to be considered; he thought that quality may be compromised, especially, in politically-driven projects, simply because the period between the decision to initiate and start up is often too short to allow for adequate quality assurance. Hsu (2009) made this idea more explicit; he posits that "when adoption is more important than implementation, decisions are frequently made with the follow-up or preparation time necessary to generate adequate materials" (74). To overcome this problem, Fullan (2007) proposes that "it is important to attempt substantial change and to do it by persistently working on multi-level meaning across the system over time" (92). By this, Fullan shows that innovation is a complicated task to carry out. In order to make it easier for innovators, it is wiser to break down complex changes into components, which can be implemented in an incremental manner.

1.5.2.3 Markee's View

Markee's (1997) framework in the area of educational innovation is seen by innovation researchers to be one of the most comprehensive models that successfully realised to

summarize and display the relevance of ideas from innovation theory. In the diffusion of innovation in language education, Markee provides Principles for language teaching professionals:

> To understand the factors that affect the design, implementation, and maintenance of innovation. His frame work is based on the questions that were posed by Cooper (1989): these include questions such as, who adopts, what, when, why, and how (118).

In terms of **who**, Markee, based on Fullan (1982), associates the **who** to the teachers. He sees that those latter are key players in the language teaching innovation. Though the teachers are different from one context to another, they tend to assume the role as implementers who carry out the process of innovation on the ground. Markee also reported that Kennedy's (1988) study shows the individuals in this process of innovation may transgress to concern often people other than the teachers, mainly those who are seen to play the role of deciders. These are individuals such as Ministry of Education, directors of schools, and general inspectors. The other part in this process are students, who are the clients. All of those individuals form the community that the innovation act targets.

In the course of the implementation process, the potential adopters, drawing from the studies by Rogers (1983) and Rogers and Shoemaker (1971), should pay attention to basic stages. Markee identifies these stages as:

1. gaining knowledge about innovation,

2. being able and persuaded of its value;

3. making a preliminary decision of whether to adopt or reject the innovation, and

4. confirming their previous decisions.

In terms of **what**, Nicholls (1983) defines "innovation as an idea, objective, or practice perceived as new by individuals, which is intended to bring about improvement in relation to desired objectives, which is fundamental and which is planned and deliberate" (cited in Markee 1993 : 231).

Chinda (2009), based on Markee's interpretation of the **what**, sees this latter:

> …what needs to be considered as a managed process of development whose principal product are teaching or teaching materials, methodological skills, and pedagogical values that are perceived as new by potential adopters or as he labelled it the who (65).

Here, Markee addresses two issues he felt were missing in Nicholls' definition. These mainly are: the notion of fundamental change and the question of whether innovations need to be planned or not (Wall, 2005 :71).

In terms of **where**, citing Cooper (1989), Markee says that "where in an innovation is implemented or is a socio-cultural, not a geographical issue" (55). Markee also stresses the importance of understanding the context where innovation takes place. In general terms, the context here refers to a social and cultural context where many factors, such as cultural, ideological and socio-linguistic are involved and currently affect it. Markee cites Kennedy (1988) who gives the name of 'sub-systems' to these factors displayed in Figure 1.1.
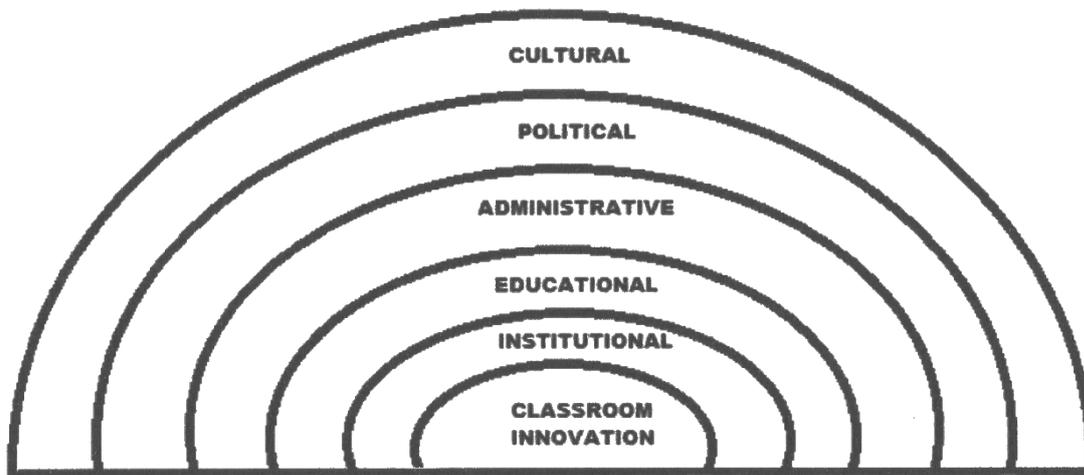
Figure 1.1: The Hierarchy of inter-relating systems in which an innovation has to operate

(Source: Kennedy, 1988 : 332, cited in Wall, 2005 : 72)

In terms of **when**, Markee discusses the rate of diffusion. He points out that this rate may vary from one type of innovation to another. He also adds that "the diffusion process tends to begin slowly and then accelerates to finally shaken" (58). Besides, Markee thinks

that innovation takes time to implement and always takes longer to implement than expected. In order to grasp this idea in the diffusion process of an innovation, it is appropriate to refer to Rogers (1995), who explained the rates of diffusion in the form of S-shaped curve (Figure 1.2). In this version, Rogers claims that most innovations follow the same pattern. First, the rate of diffusion is slow in the beginning, but then after the adoption of the rate by individuals, the rate accelerates. This is indicated by the step climb in the curve.
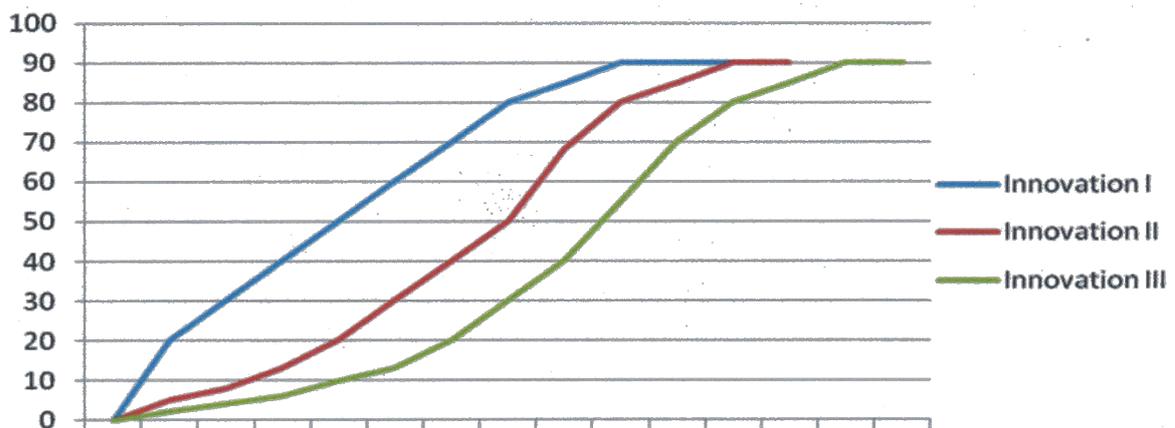
Figure 1.2: The Rate of adoption of an innovation (The S-shaped diffusion curve)

(Source: Rogers, 1995: 11, cited in Wall, 2005 : 74)

Rogers (1995) makes clear what the attributes of the rate of diffusions are. In his words, he states that:

> the rate on adoption is usually determined by five types of variables: the attributes of innovation, the type of innovation, decision, communication channels operating in the environment, the nature of the social system, and the extent of the agents (cited in Wall, 2005 : 74).

In terms of **why**, Markee discusses the characteristics of adopters and characteristics which can facilitate or hinder innovation. About the characteristics of adopters, Markee refers to Rogers (1993) and emphasizes on the assumption to consider five categories. These were discussed with some detail under the **who**. He adds that giving too much importance to the characteristics of adaptors make it possible that the agent would be more convinced about the adoption of the innovative act. On this particular point, Rogers used the phrase 'audience

segmentation' to talk about the various communication channels or appeals that are used to target different categories of adopters (Rogers, 1993, cited in Wall, 2005 : 74). The second factor that was discussed by Markee is about the features of successful innovation. Rogers (1995) stresses that it is the adopters perceptions of the features which decide whether to adopt or reject the innovation. Drawing on this assumption, Markee points out five fundamental crucial attributes to adopt or reject the innovation. These attributes were illustrated in Rogers' view (see 1.5.2.1).

Finally, in terms of **how**, Markee (1997) describes five different approaches to affecting change: the social interaction; centre-periphery model; research-development and diffusion model, problem-solving model; and, the linkage model.

1. The social interaction model: "it sees a diffusion of an innovation as a matter of communication: individuals belong to one or more network and information about spreads and colleagues interact with others in their own social grouping" (62-63).

2. The research-development, and diffusion model: "it assumes that research, long-term planning and specialist teams working on different aspects of development can ensure high-quality innovations" (63-64).

3. The centre-periphery model: "it is more 'top-down' policy-makers decision of whether and innovation will be adopted and from it should take, and pass their decision on the subordinates who must try to manage the implementation" (ibid).

4. The problem-solving model: "it is a 'bottom-up' model, where it is the potential users of innovation who decide whether there is a need for change. They identify possible, trial and evaluate them, and repeat the process until they reach satisfactory outcomes" (67-68).

5. The linkage model: "it is corporate features from the social interaction, problem-solving models, and which acknowledge that different approaches should be used in different situations, depending on the type of problem that need to be solved" (68).

1.5.2.4 Henrichsen's View

In congruence with the previous illustrated innovation views discussed earlier in this section, Henrichsen's (1989) model suggests a full understanding of the diffusion and implementation of innovation processes required not only an examination of the innovation itself, but also an in-depth examination of (a) the role of the change agent (eg. policy makers, deciders of innovation,), (b) the role of the adopted (eg. teachers and students), (c) the various stages of the innovation diffusion process (eg. decision making, adoption, implementation, diffusion), and (d) the local constraint which reformers operate. In other words, it is crucial to understand the context where innovation would occur and take part, the length of time that is required for successful innovation, and the factors that are present in the context where the innovation is expected to happen (Andrews, 2004). Without an understanding of these different components of the innovation process, innovators will find it very complicated to carry on what they want to innovate.

In this respect, from his attempt in diffusion innovations in English language teaching, Henrichsen (1989) proposes a hybrid model of the diffusion/Implementation process. The model consists of three main elements: 'antecedents', 'process', and 'consequences'.

1. The 'antecedents' component of the model focus on the significance of the set of conditions of the educational context or environment before an innovation is introduced. On this point, Henrichsen insists on that those who postulate for an innovation must be aware of the characteristics of the intended 'user system', the characteristics of the 'users', traditional pedagogical practices, and the experience of the pervious reforms before they decide on the suitable innovation to be carried out. The characteristics of the 'user system' correspond to the structure and power relationships in schools and society. The characteristics of the intended 'users' of the innovation process include the used attitudes, values, norms, and abilities. Traditional pedagogical practices consist in deriving from a

variety of cultural and historical influences. Finally, the experiences of the previous reforms will provide an understanding on how to achieve the goal or how to overcome the difficulties present in the innovation.

2. In the 'process' component of the model, Henrichsen describes and analyses the factors which stand as facilitators and/or hindrances to change; he lists the factors as follows:

- Within the innovation itself, including originality, complexity, explicitness, relative advantage, triability, observability, status, practicality, flexibility, adaptability, primacy, and form;

- Within the 'intended user system', including geographic location, centralisation of power, and administration, size of the adopting unit;

- Communication structure, gap orientation, and balance, learner factors, student factors, capacities, educational philosophy and examination;

- Within the inter-elemental, including compatibility, linkage, reward, proximity, and synergism.

3. In the 'consequences' component, the hybrid model provides different types of the innovation decisions and outcomes. In this section, Henrichsen describes how a decision to adopt or reject an innovation can be changed at a later stage; he also describes the types of innovation decisions, collective decisions, authority decisions, and contingent decisions. Besides, he labels the types of outcomes that can be in mediate or delayed manifest or latent, and functional or dysfunctional functions (Wall, 2005 : 83-86, Chinda, 2009 :66-67).

Based on this discussion, different views on innovation in education provide implications for the present study since the major aim in this research is to repair the myriads of anomalies present in the current testing system adopted by EFL teachers in the Algerian secondary schools, and hence seek to implement an ATM; besides, another implication

mainly concerns to examine the impact of the intended innovation on those who are concerned by it in the context to be explored. In this respect, this synthesis of the available literature review on educational innovation has led us to consider that Rogers (2003) model has provided a general definition of innovation, its basic characteristics, and its diffusion process. Fullan (1991) has proposed broad phases of change in innovation, as well as factors affecting each phase. Fullan, through his model, has made clear that the process of innovation should be seen as a process rather than as an event, and all the participants who are affected by this act have to find their own understanding for the change. Markee's (1997) model has yielded specific perspectives in the domain of innovation in education. The set of proposed questions have proved to be crucial in making this process true. Finally, Henrichsen's (1989) hybrid model has offered insights into the fundamental factors affecting the different stages in the implementation and diffusion of innovation.

From this synthesis of the implications of "innovations" views in this study, it is ostensible that different theories of innovation and change have provided the researcher in the present exploration useful insights on how one should proceed to implement subjects that are new for the people concerned by this change and to serve to bring positive outcomes. Knowing that not all what was discussed above by these theories can be taken for granted, it is essential to note that there is a dire need to adapt the contents of these models to the subject and objective of the present study, so that conceiving things in such a way makes these frameworks effective in their uses.

As has been pointed out at several instances in this review of the literature on innovation in education, bringing about any kind of change can be extremely long, complex, and difficult. Research into washback has consistently shown that tests, in many cases, can be seen as effective and useful levers for innovation in education. This is why the final implication of this discussion of innovation in education to the current study is that it has

enabled the researcher to base his investigation on a clear-founded theoretical background, and, second, it has also offered the understanding of how systematic the process of implementing a new testing system should be an issue which is very indispensable to diagnose the factors affecting the different stages of this innovation in this study.

1.6 A Washback Model for the Present Study

In section 1.3 in this chapter, the functions and mechanisms of washback through which it is believed to operate were investigated. The present section explores a washback model arrived at from the set of reviews of different empirical studies in various contexts. Its fundamental aim is to investigate the effects of the introduction of new tests in this research and to consider the nature of evidence required to support claims of a washback effect.

The model proposes that the nature of washback from language tests flows from overlap, the distance between the contents, and the instructional objectives set out in the relevant, taught syllabus. The greater there is a correspondence between the two, the likely positive washback becomes. Nevertheless, in this model, washback is not simply a matter of test design, it is realised through, and limited by, participant characteristics. Participants perceptions, attitudes, and reactions of test importance and difficulty, and their ability to accommodate to test demands, will moderate the strength of any effect and certainly the evaluation of its value.

To provide a structure of this model, the researcher used an adapted model of washback that is based on a framework that was suggested by Hughes (1993). It is worth noting that this latter is very common in the literature review on the most current research studies on washback. The researcher discussed this model with some detail in section 1.4. The choice of this washback model among many other available models in the language testing area is manifested by the fact that such a framework is very appropriate to the nature of the present research. In what follows, the different components that form the adopted and adapted

washback model in this study are discussed. However, before proceeding to this, it is worthwhile to mention that to describe washback phenomenon and how it occurs, it is important to distinguish between participants, process, and product, recognising that these three components may be affected by the nature of a test.

1.6.1 Washback on Participants

In considering this first component in the adopted framework in this research, one can point out that the participants' behaviour can either support or override the intended washback effect of the introduction of the new testing system. As noted by Shohamy et al. (1996), the results obtained from tests can have serious consequences for individuals as well as programmes, since many crucial decisions are made on the basis of test results. In this study, the researcher called for one of these two important sorts of washback, that is, washback on participants. This idea overlaps, to some extent with, Bachman's and Palmer's (1996 : 30-31) micro-level of washback. On this point, Bailey (1999 : 12) views that the participants either teachers or students affected by washback may be influenced by information that a test bears prior to its administration, or by 'folk-information' (such as reports from students who have taken earlier version of the test). Besides, these participants may also be influenced by several sources of feedback following the administration of the test. These would include the actual test scores provided by teachers, feedback from students, and feedback from the teachers to the students scores.

To access the participants' attitudes in a washback study, the literature review on this issue presents myriads of means of investigation. Alderson and Wall (1993) point to the inadequacies of relying on survey data in isolation, but acknowledge that surveys can help to explain teachers' behaviour by probing, understanding, and beliefs. Watanabe (2004) proposes that qualitative data methods such as interviews in understanding washback in context can provide access to the world view of the participants. He adds that qualitative

interview can also assist the researcher both in the design on more quantitative instruments and in the interpretation of the results (cited in Green, 2007 : 28).

1.6.2 Washback on Process

The second component in the adopted model is washback on process. Hughes (1993: 2) defined process "as any actions taken by the participants which may contribute to the process of learning". Specifically, he included processes such as materials development, syllabus, teaching methods, contents of teaching, learning strategies, and assessment.

To understand better washback on process, a number of language assessment researchers recommended that the triangulation of perspectives, incorporating the views of teachers is needed (Watanabe, 2004; Shohamy, 2001; Turner, 2001; Alderson & Wall, 1993). Hence, they suggest that both questionnaires responses and interview data will need to be sustained by another instrument, Cheng (1997), citing Bailey (1999) agrees that "observation allows for a richer understanding of washback than surveys alone and argues for a combination of asking through surveys and interviews - and watching through observation" (cited in Green 2007 : 29).

1.6.3 Washback on Products

The last component in the adopted washback model in the current study is washback on products. Hughes (1993) defines "the products associated washback on products as what is learnt (facts, skills, etc.) and the quality of learning". What is notable on this component is that it is sometimes difficult to untangle it from the two other components that are participants and process. For Bailey (1989), much of the literature review about participants and washback describes the various processes participants try to increase. Such processes include aspects, such as reviewing what one carries on in teaching. Shohamy (1993) highlights "processes as well when claims that negative washback often brings about under

emphasis on the means by which the learner arrives at proficiency" (186). This means to include processes and products in the model.

To access language outcomes, most of language assessment researchers claim that it is complicated to measure products. The reasons for the lack on consideration given to tests include the problem of comparing non-equivalent, often-distant groups and the selection of alternative outcome measures (Green, 2007: 29). For Madaus (1988), in evaluating outcomes, it is important to bear in mind the circularity of evaluating test impact through score gains. A rise in score does not obligatorily imply that there is an improvement in learning. Rather, the score may mask the reality that there is no positive washback of the test on the final outcomes of learners. Because of this, and in order to make washback on products actual, in the current study, the researcher would work on the evidential link between test design issues and test score interpretation. This can result in gains on the newly testing system.

Having explained the choice of a washback model, highlighted its structure and the components it comprises, introduced the appropriate research methodology to be employed to access each of these components in this research, now we turn to dependent variables of the model arrived at above. This raises how washback can be recognized and gives a clear picture on its nature. A thorough examination of these dependent variables is needed and will be presented further through the two fundamental studies: the "Preliminary Study" and the "Final Study" in Chapters 3 and 5.

On this matter, Green (2007) cites Wall and Alderson (1993) who argue that it is necessary to define dependent variables in washback research and their 15 washback hypotheses suggest predictions regarding content (**what)**, methods **(how)**, rate, sequence, degree, and depth of teaching and learning as potential dependent variables for investigation. In the same vein, calling for the same explicitness of these dependent variables, Bailey (1999) remarks that washback studies can broadly be divided into those focusing on perceptions and

57

those concerning actions. Hughes (1993), whose ideas are the starting point of the washback model adopted in this research, provides his model that attempts to tackle the dependent variables as a basis for research which encompasses both perceptions and actions and links these two variables to learning outcomes. Bailey (1996), who developed this model presents it in the form of flow diagramme, where the conditions outlined for washback are met. In this sense, washback will occur to participants, affecting their attitudes towards work. Participants' attitudes will affect processes, including both what participants do, and how they do. Processes concern aspects such as teaching materials, teaching, and learning. In their turn, these processes will influence the product: the content and the extent of learning.

Drawing on the reviewed literature, dependent variables in this study will include the effects of the prevailing, existing testing system in the Algerian secondary schools on participants' attitudes and beliefs, the content and methods of teaching and learning and the students' outcomes, in the form of their scores and self-assessed gains. Likewise, the same way of conceiving these dependent variables will be adopted with the effects of the newly, introduced testing system in the context under exploration on participants' attitudes, beliefs, and reactions to that innovation. Each of these facets to be considered poses its own challenges for the researcher in this investigation. Therefore, this study aims to explore how different components in the Algerian educational system reacted when washback was strategically anticipated to determine the possible areas of washback intensity in teaching and learning English in the Algerian Secondary schools and to define the interrelationship between the who changes, **'what'**, **'how'**, **'where'**, and **'why'**.

Conclusion

To summarize, this chapter reviewed a number of issues related to the central research concept in this research. Crucially, an attempt has been made to elucidate the origins, definitions, functions, and mechanisms of washback in language testing. In addition, this literature review on the concept has also considerably helped us to display the power and authority of tests on the teaching and learning processes, indicate how language tests become effective ways for influencing educational system, prescribe the behaviour of those who are affected by their results. Some ideas related to the question of impact of tests on teaching and learning either in positive or negative sides are still not well explained, and the raised questions remain without thorough and comprehensive answers. Next to this part, in this chapter, an array of assessment studies on washback have revealed that a large number of investigations on this phenomenon are from different perspectives and multiple levels are available. In the meantime, those studies have shown that a few of them are of empirical nature, and findings become fewer when research turns to explore the washback effects of language tests in classrooms.

The following displays the theoretical background of the research methodology employed in the present study.

Chapter Two: Research Methodology for this Study

Introduction

This chapter describes the research methodology underlying the research. It comprises four sections. Each one of these attempts to explain a particular area or a relevant issue to the research strategies and research methods of the whole research design.

The first section of the chapter describes the overall research design wherein an elucidation of matters, such as the research framework and the research methodology for this study is displayed. The second section elicits the research time that the investigation took from the initial phase until the end of the study. The next section deals with the subject of ethics in this research through which the researcher shows the basic ethical guidelines he followed to avoid any form of bias or influence that can minimize the validation and credibility of the final outcomes in this study. Following this section, the researcher's roles and his epistemological positioning are described. The aim of such a presentation is to show how the researcher proceeded to carry out the current exploration.

2.1 The Research Design of this Study: An Overview

2.1.1 Research framework

The research puts emphasis on the importance of 'context', 'setting', and 'subjects frames' (Marshall, Rossman, 1989:46). The methodology adopted in this study sought to capture the reality, variation, and the complexity of the changes in day-to-day classroom practice as well as changes within the Algerian educational context. In order to understand current practices, the researcher engaged directly in the local scene, spent sufficient time to understand how action took part in its specific context, gained access to these practices, and observed how behaviours of participants changed over time due to the innovative act subject

to the present study. Without this direct involvement, an understanding of the studied phenomenon was not possible.

Due to the nature of this newly implemented test and its impact on the educational system as a whole and the learning of English as a subject in this system, this study explored the area of teaching and learning of this subject in in-depth. The level of the washback effect on the ATM consisted in the micro-level, the level of teachers and students. The researcher suggested that observing the educational context at the micro-level was very crucial to allow one to identify gaps or mismatches between the teaching/ learning practices and the way by which learners are constantly assessed and evaluated. Accordingly, in the present case an understanding of the perception of the relevant gaps and mismatches within the educational context should greatly improve the knowledge and understanding of how and in what areas the suggested achievement tests drawn out from the ATM can actually influence the teaching/ learning process.

As has been noted in chapter 1, the washback effect is a powerful phenomenon. Indeed, it is so since it not only influences different people at different levels within educational contexts, but also this phenomenon has a strong impact on many aspects of the teaching/ learning process. This is why it is very important to gain an understanding on research findings about teaching and learning, make references to the existing models of teaching and learning, and also draw on theories in educational change, since the notion of change is the fundamental principle of the washback effect.

For Cheng (2005), classroom-centred research investigates what happens inside the classroom when the two main parties teachers and students come to interact. It also seeks to answer related questions to the issue of the process variables of the classroom and in the mean time looks for appropriate answers to questions such as how and why it is that things happen as they do in the classroom. The focus on such aspects is actually the purpose of the

current research of the washback study because it investigates how this particular phenomenon works in the educational context and in the classroom.

Accordingly, language researchers have chosen to look at the classroom as a place where language learning takes part in terms of the language input provided by the interaction between the teacher and students. In this respect, the classroom should be the first place to consider if one actually wants to understand how to help the learners learn more efficiently and how teachers and students come to interact together. Likewise, the classroom is the first place to look at if one wishes to know whether a proposed new test can bring about changes in classroom teaching and learning, and the subsequent types of changes that can occur.

Consequently, regarding the existing different approaches to classroom investigation, Chaudron (1988) summarized four main prevailing traditions in classroom research. The first tradition is referred to as the 'psychometric'. This latter involves context, passage, and product studies, which attempts to investigate the quantitative relationships between various classroom activities or behaviours and language achievement. However, this tradition has been criticized for its neglect of some existing variables in the classroom and for its try to establish a linear relationship between input and output variables via quantitative analysis (Allwright, 1988). For many language testing experts, this tradition is not adequate for studying washback as this phenomenon is very complex to have a possible linear input and output relationship.

The second tradition is the one known as 'interaction analysis'. It refers to observational schemes adopted by researchers in order to provide a description of the teacher-student in the classroom. For Moskowitz (1971), the designed observational schemes are meant to help teachers discover and evaluate their own classroom behaviours so as to become ultimately the teachers' agent for change. Like the previous first tradition, this approach was also criticized in that the used observational schemes were regarded not to be able to capture the complete

picture of classroom interaction (Bailey, 1985). Hence, for the present research, this form of analysis is considered too small to understand the complexity and multidimensional phenomenon of washback. The other tradition is seen by many researchers as a tradition that arose from a linguistic perspective, and attempts to fully analyse the discourse of classroom interaction in structural-functional linguistic terms. This third tradition is often labeled as the discourse analysis approach. In its essence, it usually views interaction as a chain of teacher and student behaviour. Its particularity is that it has the potential to be applied in a quantitative fashion. Like the previous tradition, this approach is also seen to be too detailed for the purpose of this study.

The fourth tradition is the 'ethnographic approach'. In comparison to the other major traditions, this one is considered by a great deal of researchers to be the most accepted and used approach in the domain of classroom research during the last twenty years. Its specificity is that it attempts to interpret behaviour from the perspective of the individual participant's unique understanding rather than from the observer's objective analysis. For Mehan (1979), this tradition provides "extensive empirical descriptions that approximate the knowledge of participants in a particular event, making the implicit explicit, and the invisible visible" (176). Nevertheless, this approach is often criticized for being unable to be generalized to other contexts.

What is noticeable, therefore, on the above traditions is that researchers have agreed that no one of these cited approaches and traditions could satisfactorily carry out an explanatory study of washback on teaching and learning. Cheng (2005) posits that to draw on the various strengths of these traditions, it would be better to adopt a combined research framework that takes from each of these approaches one which is suitable to the given study. On this particular assumption, Allwright (1983) notes two viewpoints on classroom centred-oriented research: one focusing on the interactive aspects of classroom behaviour and the

other focusing more on teacher talk as input. This viewpoint has brought classroom research on language teaching and learning nearer the sociological tradition of such educational researchers as Hymes (1972), and the ethnographic work of such educational researchers as Mehan (1979). This second viewpoint, focusing on teacher talk, has brought much research to language acquisition, with crucial variables as the input (Krashen, 1981). Hence, the present study puts more emphasis on these two viewpoints: interactive behaviour and teacher talk in the classroom since these two approaches are so closely related to the intended changes that the implemented ATM would make.

In terms of relevant models, the adoption of a combined framework based on the approaches discussed earlier had led the researcher to provide two teaching and learning models of classroom study at different levels and different involved contexts. The two suggested models are the ones of Chaudron (1988), Dunkin and Biddle (1974). Dunkin's and Biddle's model serves as an initial guide for the classifications of the variables and behaviours in studies of teaching. It deals primarily with the process variables of the classroom component where a teacher and students interact with each other-what teachers do and what students do in the classroom-. Concerning the teacher presage variables and students' characteristics of the context variables, they were studied through two surveys of teachers and students. Stern (1983) defines the presage variables as the characteristics, which teachers bring to teaching, their own experience, their training, and their personal qualities. The context variables concern mainly the environment and the conditions where the teachers operate. In the present study, the community and the environment were not chosen as the focus of the study. The other component, product variables, was not investigated in this study due to the nature of the research and the constraints of time. With the short span time, only limited changes could be expected in students' learning outcomes.

Chaudron's model deals with a model of four issues that concern the effectiveness of classroom interaction. As has been stated so far, interaction in the classroom is the focus of the present study. Chaudron (1988) outlines the basic issues as very crucial in improving classroom interaction. The first issue, learning from instruction, relates the basic assumptions of the acquisition of certain fundamental units or elements, the integration in functional relationships and applications, and producing and practising some operations involving these elements. The second issue, teacher talk, concerns some categories of language teaching as the speech rate, syntax, vocabulary, pragmatic functions, and others. The third issue is about the learner behaviour, linguistic behaviour, learning strategies, and social interaction. The final issue, interaction in the classroom, reveals the fundamental concepts of classroom interaction that are turn taking, questioning and answering, negotiation of meaning, in contrast to a more traditional view of teaching and learning which conceptualizes classroom interaction, as the conveyance of information from the knowledgeable teacher to 'empty' and passive learners.

In addition to these two models for research into classroom teaching and learning, the researcher considered the dire need to take into account the importance of washback functions on multiple levels. In this respect, the ideas of Fulhan (1983) and Fulhan and Stiegelbawer (1991) have been drawn upon. The authors strongly argue the need for teaching and learning to be studied within multiple perspectives. Fulhan (1983) makes this point clear enough in that he posits that the process of innovation is multidimensional. Fulhan's idea coincides with Hughes' (1993) trichotomy of backwash model that was mentioned previously in chapter 2. Hughes' trichotomy focuses on examination influences on participants, processes, and products. Bailey (1996) elaborates Hughes' trichotomy further to include different levels of influence within the trichotomy.

Drawing on all sources of research in the classroom at the micro-level, the present research proposes a working model to study the complex and multidimensional nature of washback. This model serves as theoretical and methodological guide for the current study. The proposed model is a combination of the two previously mentioned models of Chaudron (1988) and Dunkin and Biddle (1974). In addition to the fact that this model has provided theoretical guidance in what to look for teaching and learning in this study, it has also supported the researcher' beliefs and ideas about teaching and learning. After reviewing the basic of the research framework for this study, the point to lay focus on at the end for this section is the fact that such an elucidation has made it clear enough for us that washback needs to be studied in relation to the above components suggested in the exploratory model and at different levels of the educational context.

2.1.2 Research Methodology: The Present Study

Because the intention of this study was to explore the washback effect at the micro-level, focusing on perceptions, values, and situational factors in complex and varying situations of the Algerian educational context, it became apparent to the researcher that both quantitative and qualitative approaches were the most appropriate for this study. Accordingly, the current researcher decided to conduct a comparative study throughout the three phases that make-up the whole investigation. Indeed, such a way of conceiving things aimed at realizing understanding of the changes that the implementation of the ATM had brought to teaching/learning within the context under study. Thus, in the first and second phases of this research, the investigator only employed qualitative research methodology. In the final, third phase of this work, in order to find out the differences between participants' perceptions, attitudes, and opinions, the researcher needed to combine the two approaches: quantitative and qualitative. This practice was meant to help comprehend the factors that affected directly

or indirectly the process of innovation and its subsequent changes that occurred in the context of research due to the introduction of the newly testing system.

Furthermore, the research methodology for the present study was selected because it offers data collection methods which potentially enable the researcher to collect rich information for the stated purposes. Also, this methodology provides methods of analysis that are grounded in the data itself. In addition to these, the data collection was planned to gain some knowledge of the participants' perspectives if the researcher was to understand the actions and reasons of individuals at different points over the period of time during which this particular study went on. For the above reasons, therefore , a mixed-.methods approach to data collection seemed to be appropriate, since there was no one all-purpose method of capturing such a complex phenomenon, like the washback effect, as different methods have different strengths and weaknesses. Indeed, it is usually claimed that by using a range of methods one can put together a clear picture of the scene.

2.1.2.1 Research Strategies

It has been noted by many scholars that combining research methods, both quantitative and qualitative data, was employed based on the goals and circumstances of the study. As mentioned earlier, a washback study should involve what teachers talk about or think of. It is also concerned with what teachers actually do in the classroom, and what real changes have been made consciously and sub-consciously inside the classroom, both individually by the teachers and the students and also between the two parties (Cheng, 2005:70). In this context, the goals and circumstances require a multi-method methodology.

According to Cheng (ibid : 24) quantitative and qualitative approaches have their own separate strengths and weaknesses. Quantitative research involves data collection procedures that result primarily in numerical data which is then analysed by statistical methods. Typical example: survey research using a questionnaire, analysed by a computer software programme

that is designed to perform statistical operations and facilitate data analysis and is by far the most popular statistical package used by social scientists today, often referred to as SPSS. However, qualitative research involves data collection procedures that result primarily in open-ended, non-numerical data which is analysed primarily by non-statistical methods. Typical examples: interview research, with the transmitted recordings analysed by qualitative analysis. The results in a quantitative approach are generalizable but are often oversimplified and show poor ecological validity. On the other side, the qualitative approach can present a more accurate picture of reality and reveal more complexities, but it is time-consuming and the results are not generalizable (Bryman, 1992; Cohen, 1976; Halfpenny, 1979; Hammersley, Hatkinson, 1983, Patten, 1987, cited in Cheng, ibid). While we test hypotheses with a quantitative method, we have derived inferences from qualitative, conceptual considerations. The qualitative paradigm permits objectivity in the form of inter-observer agreement (Chaudron, 1986; Howe and Eisenbart, 1990; Martyn, 1992; Miles, 1979, cited in Cheng, ibid). Nevertheless, for a number of methodologists, quantitative and qualitative researches are mutually dependent. That is why, to serve the exploratory research purpose with teaching and learning variables, neither of the single approaches would be sufficient to answer thoroughly the research questions of this washback phenomenon.

Therefore, for the purpose of this thesis, the researcher adopts the framework put forward by Dörnyei (2007) who identifies eight major research strategies within the inquiry: ethnography, case studies, experiments, phenomenology, grounded theory, action research, mixed methods, and longitudinal study. In the present research, the researcher adopted the above framework, and employed five strategies: ethnography, case study, experiment, mixed-methods and longitudinal study. Below is an elucidation of the tenets of each one of these five strategies.

2.1.2.1.1 Ethnography

Ethnography is a strategy of inquiry in which the researcher studies an intact cultural group in a natural setting over a prolonged period by collecting, primarily, observational and interview data with the aim of getting an in-depth understanding of how individuals in different cultures and subcultures make sure of their lived reality (Creswell, 2009; Hesse-Biber & leavy, 2006). Ethnographic studies intend to explore the culture or shared experiences that influence by understanding the attitudes , knowledge , and beliefs that affect behaviours of the people within a community (Ladico, Spaulding and Voegtle, 2006, cited in Chinda, 2009:08-9). In this respect, ethnographers tend to be very sensitive to the matter of reflexivity and the way it affects their perceptions of the culture or events they wish to describe. What concerns them is that the conceptual tools they use to understand the culture or the events being studied are not, and can never be, neutral and passive instruments of discovery. According to Denscombe (2010:79-80) ethnography has the following characteristics:

- It requires the researcher to spend considerable time in the field among the people whose lives and culture are being studied. The ethnographer needs to share in the lives rather than observe from a position of detachment. Extended field works allow for a journey of discovery in which the explanations for what is being witnessed emerge over a period of time.

- The ethnographer is generally concerned to find out the members of the group (culture being studied understand things, the meanings they attach to happenings and the way they perceive their reality).

- Ethnography generally prefers a holistic approach which stresses processes, relationships, connections, and interdependency among the component parts.

Regarding the last characteristic, Mallinowski (1922) stresses the holistic perspective of ethnography. He notes that

> One of the first conditions of acceptable ethnographic work certainly is that it should deal with the totality of all social, cultural, and psychological aspects of the community, for they are so interwoven that no one can be understood without taking into consideration all the others (16).

Based on the above interpretations of ethnography, the purpose of this research strategy is to provide descriptions which stand in their own right without the need to worry about how representative the situation is or what the broader implication on what might be in terms of other events or cultures of the type, or of contributing to wider theories. Hence, ethnography is a strategy that aims at judging the depth of its portrayal and intricacy of its description (Chinda, ibid: 84). It is usually mentioned that ethnographic reports usually comprise a 'thick description' of the situation capturing the full complexity of the nuances in interaction, cultural, practices, and beliefs of the group under study. Nevertheless, some methodologists argue that the participants or 'key informants', on whom ethnographers depend to provide them with the richest insights into the culture, the issues observed in the study, as well as the unwritten rules of the group, might show and tell what they think other researchers want to see or hear, as well as hide things and even tell lies (Delamont, 2004:212)

In the context of the present study, ethnography could be a study of a group of teachers and/or students in their institution over a period of time, which might be a term, two terms, a school year, or more, in order to take notes, interview teachers and students, and record their answers (Richards, 2003:15). Richards (ibid) also suggests that this strategy of inquiry can provide a means of understanding teachers' own professional worlds. Therefore, in this study, the current researcher behaved as if he is a member of the staff. So, he attended teaching sessions in the field work during a period of time in about 123 hours in two successive school years (2009/2010 - 2010/2011). Also, he recorded interviews with the participants to collect

information about the way teachers teach, assess, and evaluate their learners. Likewise, he did the same with the learners.

2.1.2.1.2 Case study

A case study is a specific instance that is frequently designed to illustrate a more general principle (Nisbet, Walt, 1984:22). It is the study of an instance in action (Adelman et al., 1980). The single instance is of a bounded system. It provides a unique example of actual people in actual situations, enabling readers to understand ideas more clearly than simply by presenting them with abstract theories or principles (Cohen et al., 2005:53). For other methodologists, a case study is a strategy of inquiry of which the researcher explores in-depth one or more individual programme, process, event, or activity (Creswell, 2009:13). Based on Stake (1995), Creswell (ibid) describes cases as being bounded by the time and activity, and researchers collect detailed information using a variety of data collection procedures over a supported period of time. They can penetrate situations in ways that are not always susceptible to numerical analysis. In this connection, Robson (2002:183) remarks, case studies opt for analytic rather than statistical generalization; that is, they develop a theory which can help researchers to understand other similar cases, phenomena, or situations. Thus, for a great deal of methodologists, a case study can establish cause and effect. Indeed, one of the strengths is that they observe effects on actual contexts, recognizing that the context is a powerful determinant of both causes and effects (Cohen et al., ibid).

The main benefit of a case study approach is that the focus on one or a few instances allows the researcher to deal with the subtleties and intricacies of complex, social situations. In particular, it enables the researcher to grapple with relationships and social processes in a way that is derived to the survey approach. For Dörnyei (2007 : 54), the decision to use a case study approach is a strategic decision that relates the scale and the scope of an investigation, and it does not, at least in principle, dictate which method or methods must be used. Indeed, a

very common strength of the case study approach is that it allows the use of a variety of methods depending on the circumstances and the specific needs of the situation.

Hitchcock and Hughes (1995:22) suggest that the case study approach is particularly valuable when the research has little control over events. They consider that a case study has several hallmarks:

- It is concerned with a rich and vivid description of events relevant to the case.

- It provides a chronological narrative of event relevant to the case study.

- It blends a description of events with the analysis of them.

- It focuses on individual actors or groups of actors, and seeks to understand their perceptions of events.

- It highlights specific events that are relevant to the case.

- An attempt is made to portray the richness of the case study in writing up the report.

In ethnographic research, a case study should provide researchers with a 'thick description' of the situation to capture the full complexity and uniqueness of the use of information (Ladico, Spaulding and Voegtle, 2007:270). It typically examines the interplay of all variables in order to provide as complete an understanding of an event or situation as possible. This type of comprehensive understanding is arrived at through the process which involves an in-depth description of the entity being evaluated, the people involved in it, and the nature of the community in which it is located. 'Thick description' also involves interpreting the meaning of demographic and descriptive data such as cultural norms, community values, ingrained attitudes, and motives. In this sense, a case study is descriptive and detailed, with a narrow focus, combining subjective and objective data. Hence, it is important in a case study for events and situations to allow to speak for itself, rather to be largely interpreted, evaluated, or judged by the researcher.

In the present study, focus was on investigating EFL secondary school classes in Biskra region. The major concern was to explore the washback effect of an ATM on teaching and learning EFL in this context. The exploration sought to understand the teachers' and students' beliefs of this newly implemented testing system. Therefore, a case study approach was employed because

- It allows the provision of an in-depth, complex, and thick description of all the participants.

- It enables the use of a variety of research methods. Further, it more or less encourages the use of multiple methods in order to capture the complex reality under scrutiny.

- It fosters the use of multiple sources of data. This is, in turn, facilitates the validation of the data.

- With this approach, there is no pressure on the researcher to impose controls or to change circumstances.

- And finally, it allows the researcher to deal with the subtleties or intricacies of the complex washback phenomenon.

2.1.2.1.3 Experiments

As experiment is an empirical investigation in which investigators deliberately control and manipulate the conditions which determine the events in which they are interested, introduce an intervention, and measure the difference that it makes. The primary idea of an experiment is to isolate individual factors and observe their effects in detail. The purpose is to find out new relationships and properties associated with the subject being integrated, or to test the existing theories. In addition to this, Denscombe (2010 : 66) remarks that experiments are generally concerned with determining the cause of any changes that occur to the thing being studied. For him, it is normally enough to show that two things that occur are limited; that they always occur at the same time or they always happen in sequence.

In social sciences, the use of an experiment would appear to be restricted to those situations where it is possible to manipulate situation and impose controls on crucial variables. Laboratories are clearly designed to aid this. There is, though sometimes a dilemma here

> It is often argued that laboratory experiments are of little use because the setting in which they take place are artificial and so the results of the ensuring findings have little validity beyond the confines of the laboratory. (Bryman, 1989: 90)

In this situation, social scientists tend to rely on observing events which occur in circumstances over which they have little control. Thus, researchers in this case are on the look-out for naturally occurring and measuring the impact of isolated variables through circumstances as they happen, without imposing artificial control. The important elements displaying whether or not it is adequate to use experimentation as a ground on which the researcher rests are whether the circumstances allow researchers to observe the effect of one variable by controlling for the effect of others.

In educational research, because they cannot always control all the variables, some social scientists have turned to what they call the 'Quasi-experimental' approach as a realistic possibility. For many methodologists, this approach relaxes the probabilistic and population distinction imposed by the other types of research methodologies, 'true' experimental design by shifting emphasis from 'cause-effect' in temporal priority to 'association' between variables. That is, quasi-experimental approaches are able to employ something approaching true experimental design in which they have control over what Campbell and Stanley (1963) refer to as 'the when and to whom of measurement' but lack control over the 'when and to whom of exposure' or the randomization of exposures- essential if true experimentation is to take place. Kerlinger (1970) refers to these situations as 'compromise design', an apt description when applied to much educational research where the random selection or random assignment of schools and classroom is impracticable. Another advantage of the

quasi-experimental approach is that it is frequently sustained by statistical analysis. The latter permits the provision of validity to the result at the end of the investigation.

In the present study, because the researcher's aim was to implement an ATM that attempts to repair the myriads of anomalies present in the Current Testing System used by EFL secondary education teachers in Biskra region to evaluate and assess their learners' progress and achievement in accordance with the instructional objectives, the quasi-experimental approach was adopted to corroborate the results obtained by the other data collection methods, and hence make the final outcomes more valid.

2.1.2.1.4 Mixed –methods

Mixed-methods apply to research that combines alternative approaches within a single research project. It refers to a research strategy that crosses the boundaries of conventional paradigms of research of deliberately combining methods drawn from different traditions with different underlying assumptions. A mixed-method strategy is one that uses both quantitative and qualitative methods. As an approach, with a recognized name and research credibility, it has only come to the fore in recent years supported by writers such as Cresswell (2009), Creswell and Planoclark (2007), Greene (2007), Johnson and Onwvegbuzie (2004) and Tashakori and Teddlie (2008). It is notable that a number of different names have been given to mixed-methods strategies: 'Mixed-methodology', 'multi-strategy research', 'investigated methods', 'multi-methods research', and 'triangulation' are just some of the alternatives. These names bear testimony to the variety of ways in which research can be mixed and the many aspects of research process can be involved. The most common form of mix, however, involves a mix of methods and, in practice, the term mixed- methods has come to be used in a general sense to cover the spectrum of ways in which mixing can occur in the research process. In effect, it has given its name to a broad approach to research (cited in Denscombe, 2010:137-138).

In social sciences, the mixed-methods approach has been endorsed by some of the most influential methodologies. Miles and Huberman (1994:319) summarize the recent emerging tradition and say:

Entertain mixed-model. We have thought to make a virtue of avoiding polarization, polemics, and life at the extremes. Quantitative and qualitative inquiry can support and inform each other. Narratives and variable-driver analyses need to interpenetrate and inform each other. Realists and idealists, and critical theorists can do better by incorporating other ideas then by remaining pure. Think of it as hybrid vigour.

Similarly on the mixed-methods approach, Dörnyei (2007:44) refers to Strauss and Corbin's (1998:34) to explicitly illustrate the tenets of this method:

Qualitative and quantitative focus of research both have roles to play in theorizing. The issue is not whether to foster the development of theory. Although most researchers tend to use qualitative and quantitative methods in supplementary or complementary focus, what we are advocating is time interplay between the two. Qualitative should direct the quantitative feedback into the qualitative in a circular, but at the same time evolving, process with each method contributing to the theory in ways that only each can.

According to Denscombe (2010 : 138-139), the mixed- methods approach has three characteristic features that set it apart from other strategies from social research. These can be summarized as follows

1. Use of qualitative and quantitative approach within a single research project: at the heart of the approach is the idea that researchers can bring together within a single research project certain elements that have conventionally been treated as either or 'option'. In most cases, the description is drawn between ' quantitative' and 'qualitative' with researchers variously writing about using both quantitative, and qualitative methods, quantitative and qualitative data, or qualitative and quantitative research.

2. Explicit focus on the link between approaches (triangulation): the mixed-methods approach analyses the need to explain why the alternative approaches are beneficial and the alternatives are to be brought together. Particular attention is given to the design of

mixed-methods research and especially the role of triangulation is justifying the use of the alternative approaches.

3.Emphasis on particular approaches to research problem: the mixed-methods approach is 'problem- driven' in the sense that it treats the research problem as the overriding concern. Other approaches, of course, share a concern for practical solutions to real-world problems. However, advocates of the mixed- methods approach regard it as the crucial driving force behind decisions about which methods to use. This means that they are prepared to use methods that come from different philosophical traditions, provided their use procedures findings that are of practical value for addressing the research problem. The mixed –methods approach in this sense, challenges the premise that social research should drive for consistency between the various components of the various research process in terms of their underlying assumptions about the nature of the social world and the possibilities this opens for the research.

In the present research, drawing upon the methodological points described above, the researcher opted for a mixed-methods approach because he thought that mixing methods can be a valuable research strategy for the bias in research methods, and the development of research instruments. The use of more than one method can enhance the outcomes of the research by providing a further and more complete picture of the thing that is being studied.

2.1.2.1.5 Longitudinal study

In the literature on longitudinal research, many methodologists consider this research strategy to be very recent. Thomson, Plumridge and Holland (2003) confirm this information, and eventually recognize that "longitudinal research as 'a primary new methodology' which is yet taking place without a relevant literature to inform and debate the epistemological or practical decisions [they] were making" (185). What is more significant about this term is the fact that still there does not exist one precise definition of what longitudinal research means.

Rather, the available and agreed upon definition claims that longitudinal study refers to a family of methods that share something in common: information is gathered about the target of the research during a period of points in time. While this explanation of what longitudinal study is seems somehow straightforward, some methodologists view it to be inaccurate since there may be a single data gathering session that will yield longitudinal information about change over time. Accordingly, several schools point out that there is not necessarily a one-to-one correspondence between the design of the study and the type of the data to be collected (Menard, 2002; Taris, 2000). Thus, for the advocates of this understanding longitudinal research must be defined in terms of both the data and the design that are used in the research. According to Dörnyei (2007:79) who referred to Menard (2002), a longitudinal investigation is research in which (a) data are collected for two or more distinct periods; (b) the subject or cases analysed are the same or comparable from one period to the next; and (c) the analysis involves some comparison of data between periods. As he concludes "the above minimum, any truly longitudinal design would permit the measurement of differences or change in a variable from one period to another" (2).

In terms of types, there are as many as four types of longitudinal design; the first type is 'prospective longitudinal studies' also known as 'panel studies'. This design implies that successive resources are taken at different points in time from the same respondents. It has also been stated that the popularity of this type of longitudinal study is because such a research strategy allows the researcher to collect information about changes at the micro-level as it really happens. Many methodologists postulate that this first type, indeed, offers a powerful non experimental method for examining development and causality. But, the same people believe that such a strategy presents some problems in that it is seen to be time-consuming and expensive to run, and it also requires a committed research team maintained over years. The second type is the 'repeated cross-sectional studies' or 'trend studies'. With

this strategy, obtaining information about change is usually by administering repeated questionnaire surveys to different samples of respondents. If the subsequent waves examine samples that are representative of the same population, then the results can be seen to carry longitudinal information at the aggregate level. Furthermore, the design makes it possible to investigate and compare changes in various subsamples.

The main advantage of this type is that it does not suffer from attrition or condition, and each wave can be made representative of the population. Unfortunately, this second type of longitudinal research presents, too, some inadequacies in that the macro perspective which is characteristic of the cross-sectional design which is not particularly suited to resolve issues of casual order or to study developmental patterns (Taris, 2000). The third type of longitudinal research concerns what is named 'retrospective studies'. What is notable with this third type is that the data are gathered during a single investigation in which respondents are asked to think back and answer questions about the past. This way of doing things is seen to be a good idea since it enables the researcher to save time and money. Once again, this other type has also shown some shortcomings in that past retrospective research has revealed the quality of the collected data can be uneven, which means that about as much as the half of the responses may be inaccurate or incorrect in some way (ibid). The last research design in this list of four types of longitudinal study is the 'simultaneous cross-sectional study'. What is significant with this type is the fact that it is partially longitudinal, because it does not involve the examination of the change over time, but rather cross age groups. In other words, this means that cross-sectional survey is conducted with different age group sampled. This design is straightforward and economical; and, because it yields data about changes across age groups, it can be used to examine developmental issues. However, as the previous types, the 'simultaneous cross-sectional study' presents a clear drawback in that it may measure

different cohorts and the observed changes may not be due to the age difference but to the special experiences of cohorts (Dörnyei, 2007 : 79-85).

In the present study, the longitudinal design employed can be classified as a 'prospective longitudinal study' of which data are gathered at different points in time from the same respondents. This type of longitudinal design was utilised because, according to Dörnyei, it offers a complex and true reflection life story of an individual participant, which was the major aim of the present research.

2.1.2.2 Data Collection Methods

As discussed in chapter 1, washback study is seen as a multivariate or multifactor endeavour that required the researcher to think of, or address more than one factor at a time. Consequently, for the data collection methods, there is ample evidence that the use of a combined quantitative and qualitative research design would certainly produce a valuable tool for the subsequent research strategy to obtain the needed data and explicit these collected information to enable the researcher to give an account of what is happening in the context. As noted in section 2.1, this type of mixed-method strategy that provides for cross-examination mechanisms is often referred to as triangulation.

Some methodologists argue on that the notion of triangulation was first used in the field of geometry and the geometric laws associated with triangles. Denscombe (2010:346) remarks that there are many applications of these laws but one extremely valuable application concerns the ability to find the exact location of a point if it is viewed from two other known positions. He goes on to add that the combination of information about angles and distances from these two other positions has been used extremely, for example, to help navigation and surveying.

Social researchers have adopted the principle of triangulation and applied it in a variety of ways. Denscombe (ibid : 38) observes that triangulation in the social context does not

equate literally with the use of triangulation by the geologists, surveyors, navigators, and others concerned with measuring exact locations of things in the physical world. Rather, he posits that this term is used in a metaphorical sense and he justifies this by the fact that social researchers are not able to make use of fixed, objective, positions, universally agreed, from which to make their observations not in the same way surveyors, engineers, and others rely on absolute, objective positions in the physical world. What a social scientist has adopted, however, is the principle that viewing something from more than one viewpoint allows the researcher to get a better knowledge of it.

In educational research, according to Denzin (1978), triangulation refers to the generation of multiple perspectives on a phenomenon by using a variety of data sources, investigators, theories, or research methods with the purpose of corroborating an overall interpretation (Dörnyei, ibid: 165). Denscombe (2010) confirms this interpretation of the term and what it implies. He notes that this approach involves the practice of viewing things from one perspective. In other words, this means that the triangulation is the different methods, different sources of data, or different researchers within the study.

What is notable with these descriptions is that all research methodologists have come up with the same point that triangulation refers to the use of more than one approach / method to the investigation of a research question in order to enhance the confidence in the ensuring findings. Thus, since this research strategy can provide complementary data and tackle an issue from different angles, triangulation is seen to be an effective data collection method that can ensure the best research and accurate measurement of things, and facts, and evidence about the subject under exploration.

Building on the work of Denzin (ibid), Denscombe (ibid) has adopted the principles of triangulation and applied it to a number of forms. In this respect, he distinguishes five major types of triangulation:

1. **Methodological triangulation** (between-methods), which refers to the use of alternative methods. This particular type allows the findings from one method to be contrasted with the findings of another. For the sustainers of this type of triangulation, there are many benefits

   • Findings can be corroborated or questioned by comparing the data produced by different methods.

   • Findings can be complemented by adding something new and different from one method to another.

2. **Methodological triangulation** (within –methods), this refers to comparison using singular methods. It can provide a check on the accuracy of the findings. This type of triangulation was first used by Campbell and Fiske (1959) and Webb et al.(1996). The use of widely contrasting methods would increase the possibility that any difference in the findings might reflect the difference in the methods rather than the actual factor that is being checked (Mc fee, 1992, cited in Denscombe, ibid, 343).

3. **Data triangulation** (use of contrasting sources of information), which entails gathering data from several sources of information. This can mean comparing data from the different informants (informant triangulation) or using data collected at different times (time triangulation). This type also concerns the idea of space triangulation (the use of more than cultural, social, or geographical context).

4. **Investigator triangulation**, which refers to the use of more than one researcher in the field work to collect and analyse data. The benefit of this type of triangulation is that it permits the avoidance of any form of bias arising from the influence of any specific researcher.

5. **Theory triangulation**, which involves the use of more than one theoretical position in relation to the data.

A broad generalization in research methodology is by and large true in that the main strengths of triangulation in social research are that by choosing more than one method for data collection, this allows the researcher to look at the research topic from a variety of perspectives, as a means of comparison and contrast, with each method approaching the topic from a different angle. Besides, this use of triangulation focuses on the validation of the findings in terms of their accuracy and authenticity. The alternative methods are used as confirmation. The other benefit of the triangulation approach is that it focuses on producing complementary data that enhances the fuller picture of the findings. In this sense, triangulation is used to complement information from other sources.

Accordingly, in order to complement each other and compensate for an individual method's weakness, triangulation is employed in the present study. The adopted methods are questionnaire, interviewing, classroom observation, and focus groups. Both 'Between-methods' and 'Data triangulation' are used to examine the methods and data. The triangulation involved is set in Figure 2.1.

Questionnaires	
• Questionnaire for teachers(01) • Questionnaire for teachers(02)	• Questionnaire for students (01) • Questionnaire for students(02)

Classroom Observation
• EFL classes of teachers in the secondary school in Biskra region who teach English under the ATM.

Interviews
• The teachers who adopted the ATM • The students. • General inspectors of Biskra region

Focus Groups
• The students

Figure 2.1: Data and method triangulation for the study.

*Arrows are used to indicate the direction of triangulation.

2.1.2.2.1 Questionnaires

Questionnaires are written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting from among existing answers (Brown, 2001:06). This data collection method is a useful instrument not only for collecting information, but also for providing respondents with structured, often numerical and analysing data in a straightforward way (Wilson and Mclean, 1994). The popularity of questionnaires is usually due to the fact that this data gathering tool is easy to administer, since it is extremely versatile and uniquely comparable of gathering a large amount of information quickly in a form that is readily processable. Indeed, because of these criteria, questionnaires are regarded as the most often employed data collection devices in both social and educational research domains.

What is remarkable about the strategies a researcher needs to follow when approaching a planning of questionnaires during the period that often precedes their realization is the fact that there exist two major approaches on the available literature on this matter. First, the advocates of the first approach view that in order to design and develop a questionnaire, the researcher has to follow three systematic stages. The initial stage concerns mainly developing a chart technique to plan the sequences of questions which is usually recommended at the very beginning. This initial stage is very helpful for the continuation of this process since the researcher is often able to anticipate the type and range of resources their questions are likely to elicit. Next, the second stage of operationalization deals with the issue of taking a general purpose or a set of purposes and turn these into concrete and researchable fields about which data can be gathered. In this sense, a questionnaire general purposes must be classified and then translated into a concrete, specific aim or a set of aims. Having decided and specified the primary objective of the questionnaire, the following step of planning involves the identification and itemization of subsidiary topics that relate to its central purpose. Third, the

final stage in planning questionnaires follows the identification and itemization of subsidiary topics and involves formulating specific information requirements, relating to each of these issues. What we have here is a move from a generalized interest or purpose to a very specific set of features about which direct data can be gathered (Cohen et al., 2005:246).

Nevertheless, the sustainers of the second approach to planning questionnaires suggest an alternative strategy to follow. In essence, this view often consists in primarily identifying the research problem, then, clarifying the relevant concepts or constructs, and finally, identifying what kinds of measures or empirical indication there are some of these. This means that in this process there is a need to pay attention to the kinds of data that are usually required to give the researcher relevant evidence about the concepts and constructs. Examples of this is to display the constructs' presence, identify, main features, and dimensions (Wilson and Mclean, ibid, 8-9)

What is significant with these two approaches to making questionnaires ready for use is the fact that both of them recognize that there is a need that the questionnaire: (a) is clear of its purpose; (b) is clear on what needs to be included or covered in it in order to meet the needs;(c) asks the most appropriate kinds of questions; (e) elicits the most appropriate kinds of data to answer the research purposes and sub-questions; (f) asks for empirical data.

In terms of types of questionnaires, there are three major ones: structured, unstructured, and semi-structured questionnaires. First, structured questionnaires are this type of questionnaires composed of most closed questions. Those latter are seen to be useful in that they can generate frequencies of responses amenable to statistical treatment and analysis. They also enable the comparisons to be made across groups in the sample (Oppenheim, 1992: 115). However, this type of questionnaires is regarded by several methodologists to be heavy on time in the research. Second, unlike the structured questionnaires, the unstructured questionnaires are based more on word-based and open-ended questions. They are considered

to be more appropriate because they can capture the specificity of a particular situation. The main advantage of this second type of questionnaires is that it can enable a rapid analysis of the data since the researcher knows the response categories, the nature of the data, and the statistics to be used. In comparison to the first types, the semi-structured questionnaires are based on a series of questions, statements, or items that are presented and the respondent is asked to answer, or comment on them in a way that s/he thinks best. In this way, there is a clear structure, sequence, focus, but the format is open-ended, enabling the respondent to respond in his/ her own terms. The semi-structured questionnaires set the agenda but do not presuppose the nature of the purpose.

The main attraction of questionnaires is their unprecedented efficiency in terms of (a) economy, in the sense they can supply considerable amount of research' data for a relatively low cost in terms of materials, money, and time; (b) simplicity to be arranged; in that it is actually for a researcher to arrange a questionnaire easier than with, for example, personal interviews. All what a researcher needs is to submit, mail, or even email the questionnaires for the respondents, who can hand back these documents using the same means; (c) standardized answers; (d) pre-coded answers, in that this allows the speedy collection and analysis of data of the researcher. In addition, these cited examples of advantages of questionnaires are not the only ones; rather, there exist some others, for example, questionnaires are also very versatile, which means that they can be used successfully with a variety of people in a variety of situations targeting a variety of topics. As a result, the vast majority of research projects in social sciences and educational research spheres involve one stage or another collecting some sort of data (Guilham, 2000; Dörnyei, 2007; Denscombe, 2010).

Since the primary aim of the present research is to explore teachers' and students' reactions to, and perceptions of implementing the ATM by the introduction of the

achievement tests drawn out from this modified, testing system , questionnaires would be an appropriate data collection method in order to allow a great deal of teachers and students to voice their opinions. Questionnaires could also be used to indicate where a more focused investigation should take place at the later stages of the study.

2.1.2.2.2 Interviews

In general, an interview marks a move away from seeing subjects as easily manilapable and data as somehow external to individuals that are usually expected to generate knowledge often through conversation (Kvale, 1996). Regarding an interview in such a way, it is assumed that this method is seen to favour exchange and interchange of views between two or more people on a topic of natural interest, with a central concern to emphasize the importance of human interaction for knowledge production and social situatedness of research data. As such, the interview is not exclusively either subjective or objective, it is inter-subjective (Laing, 1967). In other words, this simply implies that interviews are useful data collection methods that often enable participants -be they interviewers or interviewees- to discuss their interpretation of the world in which they live, and to express how they regard situations from their point of view. In these interpretations, the interview is not only concerned with collecting data about life: it is part of itself, its human embeddedness is inescapable (Chinda, 2009:144). Converging with this view, Rapley (2004) notes that "interviews are social encounters where speakers collaborate in producing retrospective (and) prospective accounts a version of their past (future) actions, experiences, feelings and thoughts" (16). In the same vein, Hesser-Bibber and Leavy (2006) add that "this method is useful when the researcher has a particular topic he or she wants to focus on to gain information from individuals" (20). They also stress that interviews, conceived in this way, are a 'measuring-making' and 'knowledge-producing' conversation that frequently occur between two parties.

87

Nonetheless, it is agreed upon among researchers that interviews are not the same as conversations; and the latter can never stand for the same uses and purposes of the former. Following this assumption, Denscombe (2010) asserts that an interview is different from a conversation. For him, interviews involve a set of ideas and understandings about the situation, which are normally not associated with a causal conversation. More interestingly, he makes his interpretation clear enough and comprehensive in that he states that when someone agrees on to have part in a research interview:

•There is consent to take part. From the researcher's point of view, this is particularly important in relation to research ethics. The interview is not done by secret recording of discussion or the use of casual conversations as research data. Rather, it is openly a meeting intended to produce material that will be used for research purposes and the interviewee understands this and agrees to it.

• Interviewee's words can be treated as 'on the record' and 'for the record'. It is, of course, possible for interviewees to stipulate that their words are not to be attributed to them, or not to be made publicly available. The point is, though, that unless interviewees specify to the contrary, the interview talk is 'on record' and 'for record'.

•The agenda for discussion is set by the researcher. Although, the degree of control exercised by the researcher will vary according to the style of the interviewing, there is a tacit agreement built into the notion of being interviewed that the proceedings and the agenda for the discussion will be controlled by the researcher (Ibid: 172 - 3).

Denscombe culminates his description of the distinction between an interview and a conversation by yielding the underlying assumptions of the systematic nature of the interviews in comparison to conversation. In doing so, he synthesizes this in these words:

The superficial similarity between an interview and a conversation can generate an illusion of simplicity. We all have conversations and it is likely that most of us do not have too much difficulty with them. So an interview should be fairly straightforward. As long as we know to whom we are going to talk and what we want to ask, the rest should be plain sailing. The superficial similarity can encourage a relaxed attitude to the planning, preparation, and the conduct of the method that would be unlikely were it to involve questionnaires or experiment. In reality, interviewing is not an easy option. It is fraught with hidden dangers and can fail miserably unless there is good planning, proper preparation and sensitivity to the complex nature of interaction during the interview itself (173).

At another level, the research interview may have a set of purposes, each of which involves a situation where the retrospective roles of the interviewer and interviewee may vary and the motives for taking part may differ; a common dominator is the transaction that takes place between seeking information on the part of one, and supplying information on the other part of the other. In a very comprehensive illustration, Cohen et al. (2005:268) propose that an interview may serve three basic purposes. First, it may be used as the principal means of gathering information having direct bearing on the research objectives. As Tuckman (1972) describes it, by providing access to what is inside a person's head, [it] makes it possible to measure what a person knows (knowledge or information), what a person likes or dislikes (values and preferences), and what a person thinks (attitudes or beliefs). Second, it may be used to test hypotheses on to suggest new ones; or as an explanatory device to help identify variables and relationships. The least but not the last, as a third purpose, the interview may be used in conjunction with other data collection methods in a research understanding. In this connection, Kerlinger (1970) suggests that it might be used to follow up unexpected results for example or to validate other methods, or to go deeper into the motivations of respondents and their reasons for responding as they do.

Concerning the types of this method, three common types are considered as the ones that are mostly used. These include examples such as the structured, semi-structured, and the unstructured interviews. First, what is notable with the structured interviews is that the interviewer follows a prepared and a very elaborate 'Interview Schedule' or 'Interview

Guide', which contains a list of the question-items to be asked to the interviewee. The advantage of this particular type of interviews is that it ensures that the interviews often focus on the central topic of discussion and that the interview covers a well-defined domain, which makes possible that the questions could be compared to one another across the times devoted to the interview. In this respect, a number of researchers stipulate that such a type of interviews resembles to a great extent to the questionnaire method than the other types of interviews. Nonetheless, the same researchers often claim that in structured interviews, there is generally little room for variation or spontaneity in the respondents' answers because the interviewee is expected to record the answers according to a coding scheme. Besides, there is also very little flexibility in the way question-items are asked by asking a standardized format it is hoped on the part of the researcher that nothing will be left to chance.

Second, with semi-structured interviews, the interviewer still has a clear list of issues to be asked and questions to be answered. This type of interviews is suitable for cases and situations when the researcher has a good enough overview of the phenomenon or the domain in question and is able to develop both questions about the topic in advance but does not want to use ready-made response categories that would limit the depth and breadth of the respondents' story. This format needs an 'Interview Guide' which has to be made and pilot in advance. With the semi-structured interviews, the interviewer will usually ask the same questions of the participants, although not necessarily in the same order or wording, and would supplement the main questions with various probes. Third, the unstructured interviews, also referred to as the ethnographic interviews, are about the type of interviews which is seen to allow maximum flexibility to follow the interviewee in unpredictable directions, with only minimum interference from the research agenda. The intension is to create a relaxed atmosphere in which the respondent may reveal more than he/she would do in formal context, with the interviewer assuming a listening role. With the unstructured interviews, there is no

detailed guide which is prepared in advance. A particular point to be raised here is that during the interview, the researcher may ask an occasional question for clarification and may give reinforcement feedback as any good communication partner which keeps the interviewer moving, but interpretations are kept to a maximum (Dörnyei, 2007). In contrast to the structured interviews, the unstructured interviews allow the interviewee to use his/her own words and develop his/her thoughts. Moser and kalton (1977) describe this specific idea. and put out that:

> The informant is encouraged to talk about the subject under investigation and the cause of the interview is mainly guided by him. There are no set questions, and usually no predetermined framework for recorded answers. The interviewer confines himself to elucidating doubtful points, to rephrasing the respondents' answers and to probing generally. It is an approach especially to be recommended when complex attitudes are involved and when one's knowledge of them is still in vague and unstructured form (cited in Cohen, 2011:273).

An important issue to raise when time comes to think of or decide to adopt this data collection method is to look for the procedures that are regarded to be crucial to make an interview ready for use. Reference here is to the process of planning interviews. In one of the most comprehensive illustrations, Cohen et al (2005) cited kvale (1996) who set out seven stages in an interview investigation that can be used to plan this type of research. In precise elucidation, the seven stages include the following:

Stage 1: Thematizing.
It is the preliminary stage of an interview study. This mainly concerns formulating the purpose of an investigation and describing the concept of the topic to be investigated before the interviews begin.

Stage 2: Designing.
This involves translating the research objectives into the questions that will make-up the main body of the schedule. What is significant in this stage is that the

interviewer has to be sure about the questions that adequately reflect what he/she is trying to find out.

Stage 3: Interviewing.
It deals with the process of setting up and conducting the interview.

Stage 4: Transcribing.
It is regarded by researchers as an important step in the process of interviewing since it prepares the required material for further analysis. The specificity of this stage is that the researcher transfers the recorded data into written texts.

Stage 5: Analysing.
On the basis of the purpose of the topic under investigation, the method for analysis is going to be elaborated. In this stage, the collected data is supposed to be interpreted in order to draw out useful outcomes to confirm or disconfirm the set of hypotheses that the research underlies.

Stage 6: Verifying;
It is to ascertain the reliability and validity of the interviews. In other words, this simply means validating the interviews findings.

Stage 7: Reporting;
It is to communicate the outcomes obtained via the interviews. The nature of the reporting itself will be decided to some extent by the interview. For instance, a structured interview will provide statistical and numerical data, while a semi-structured interview will yield word-based accounts that take up considerably more space.

One major issue here is that validation must take place at all the seven stages of the interview investigation. This is a crucial question facing the researcher who employs this data collection method. This process mainly concerns checks to see if the collected information is broadly confirmed by other people and other sources. Hence, the interviewer

must be at pains to conduct the interview carefully and sensitively. It is also important for the interviewer to keep upmost in one's mind the fact that the interview is a social, interpersonal encounter, nor merely a data collection exercise.

For applied linguists, the main advantage of interview research is that it can capture information that would not be collected through other data collection methods such as, questionnaires, and classroom observations. This is true because interviewing is certainly the genre which is usually welcomed by people and is a socially acceptable way that most of the questioned people feel comfortable with and which can be used in a variety of situations and focusing on myriads of topics to yield in-depth data. On this last assumption, Denscombe (2010: 192-3) counts a series of advantages of interviews. Some on these benefits are as follows:

- Interviews are particularly good at producing data which deal with topics in in-depth and detail. In this sense, subjects can be probed, issues pursued, and lines of investigation followed over a relatively lengthy period.

- Interviews require only simple equipment and build on conversation skills which researcher already have.

- Interviews are a good method for producing data based on respondents' priorities, opinions, and ideas. In this way, respondents have the opportunity to expand their ideas, explain their views, and identify what they regard as the crucial factors.

- As data collection methods, interviews are probably the most flexible. Adjustment to the lines of enquiry can be made during the interview itself. Interviewing itself allows for a developing line of enquiry.

- Direct contact at the point of the interview means that data can be checked for accuracy and relevance, as they are collected.

- Compared with other data collection methods such as, questionnaires, observation, and experience, with interviews, there is a more personal element to the method, and people tend to enjoy the rather rare chance to talk about their ideas at length to a person whose purpose is to listen and note the ideas without being critical.

In the present study, realizing the difficulty to opt for only one data collection method, of interview investigation was due to the assumption that such a method can help to overcome the weakness of one or the other instruments. In the mean time, it can also strengthen the positive side that interviews underlie. In particular, the present researcher chose to employ the semi-structured type of interviews. The latter is thought of to be the one that can satisfy to a great extent the exigencies of the present investigation. In this context, semi-structured interviews offer a possibility to gain useful information on the respondents' perceptions, attitudes, and reactions to both the Current Testing System and the ATM in the Algerian secondary schools.

2.1.2.2.3 Classroom Observation

Observation is another distinct way of collecting data. This data collection method is usually seen to offer the researcher the opportunity to gather 'live data' from the 'situations'. Traditionally, observation has been characterized as non-interventionist since the researcher does not seek to manipulate or alter the situation or the subjects; he does not pose questions for the subjects, nor does he deliberately create 'new provocations' (Alder & Alder, 1994). Moreover, because observed phenomena are less predictable, there is a certain adherence to this form of data collection that is denied in other forms of research such as questionnaires or tests. On this point, Denscombe (2010) advocates that:

observation does not rely on what people say they do, or what they say they think. It is more straightforward than this. Instead, it draws on the direct evidence of the eye to witness events at first hand. It is based on the premise that, for certain purposes, it is best to observe what actually happens" (196).

From a specific research perspective, observation is fundamentally different from questioning because this data collection method provides direct information rather than self-report accounts, and thus it is one of the three basic data sources for empirical research (Dörnyei, 2007). In more practical concerns, this enables the researcher to understand the context of programmes, to be open-ended and inductive, to see things that might otherwise be consciously missed, to discover things that participants might not freely talk about in interview situations, to move beyond perception-based data, and to access personal knowledge (Cohen et al., 2005). Morrison (1993) converges with the above and notes that observation can aid the researcher to gather data on the following:

- The physical setting (eg. The physical environment and its organisation);

- The human setting (eg. The organisation of the people, the characteristics and make-up of the groups or individuals being observed, for instance gender, class, …);

- The interactional setting (eg. The interactions that are taking place, formal, informal, verbal, non-verbal, etc.); and

- The programme setting (e.g. the resources and their organisation, pedagogic styles, curricula and their organisation) within the sphere of classroom research, and as highlighted in 2.1.2, the review of literature on classroom observation in educational settings has displayed that this data collection method is fundamental in understanding what is happening in a particular place. "classroom observation views the classroom as a place where interactions of various kinds take place, affording learners opportunities to acquire L2" (Ellis, 1994: 465). Allwright sees "interaction as the

95

fundamental fact of classroom pedagogy"; "everything that happens in the classroom happens through a process of live person-to-person interaction" (156). This perspective has led researchers to observe and describe the interactional events that take place in a classroom in order to understand how learning opportunities are created (cited in Cheng, 2005:92).

As we saw with the previous data collection methods, with observation, there are many ways how applied linguists categorize the different types of this instrument. As a first example, Denscombe (2010) considers that there are two kinds of observation research used in the social sciences. The first kind of these is the systematic observation. Systematic observation has its origin in social psychology - in particular, the study of interaction in settings such as school classrooms. It is normally linked with the production of quantitative data and the use of statistical analysis. The second kind is the one referred to as participant observation. This is mainly associated with sociology and anthropology and is used by researchers to infiltrate situations, sometimes as undercover operation, to understand the culture and processes of the group being investigated. It is normally associated with qualitative data. For Denscombe, even though these two kinds might seem poles apart in terms of origins and use in current social sciences, but still they share some vital characteristics. Some of these are as follows:

- **Direct observation**: the obvious connection is that they both rely on direct observation. In this respect, they stand together, in contrast to methods such as questionnaires and interviews, which base their data in what informants tell the researcher, and in contrast to documents where the researcher tends to be one step removed from the action.
- **Fieldwork**: the second common factor is their dedication to collecting data in real-life situations, out there in the field. In their distinct ways, they both

involve fieldwork. The dedication to fieldwork immediately identifies observation as an empirical method for data collection. As a method, it requires the researcher to go in search of information at first hand, rather than relying on secondary sources.

- **Natural settings**: field observation-distinct from laboratory observation-occurs in situation which would have occurred whether or not the research had taken place. The whole point is to observe things as they normally happen, rather than happen under artificially created conditions as laboratory experiments. There is a major concern to avoid disrupting the naturalness of the setting when undertaking the research. In this approach to social research, it comes to minimize the extent to which the presence of the researcher might alter the situation being researched.

- **The issue of perception**: it is usually claimed that the researcher's perception of situations might be influenced by personal factors and that the data thus might be unreliable. They tend to offer very different ways of overcoming this, but both see it as a problem that needs to be addressed.

In a second example, Dörnyei (2007) proposes two dichotomies: 'participant' versus 'non-participant observation' and 'structured' versus 'non-structured or unstructured' observation through which one can observe classrooms. At the level of the first dichotomy, it is thought that the 'participant' observer becomes a full member of the group, taking part in all the activities. This is the kind of observation that usually takes part in ethnography studies. In other contexts, however, the researcher can be minimally involved in the setting and if he/she behaves in such a way, the researcher can be described as a 'non-participant' observer.

Hammersley and Atkinson (1983:93-5) suggest that comparative involvement may come in the forms of 'participant' observer and the 'non-participant' observer, with a degree

of subjectivity and sympathy, whilst comparative detachment may come in the forms of the observer as participant and the complete observer - where subjectivity and sympathy are key characteristics. Both complete participation and complete detachment are as limiting as each other. Moreover, about the second dichotomy the distinction between 'structure' and 'unstructured' observation is similar to the 'quantitative' versus 'qualitative' distinction in observational terms. Highly structured observation involves going into the classroom with a specific focus and with concrete observation categories, whereas unstructured observation is less clear on what it is looking for and the researcher needs to observe first what is taking place before deciding on its significance for the research (Cohen et al., 2005). The former approach involves completing a scheme, while the latter deals with completing narrative field notes, often supplemented by maps or diagrams. This is, of course, a continuum, and in practice usually some combination of the two approaches takes place.

A significant problem in observation has been pinpointed by researchers in that it has been shown that facts recorded by one researcher are very likely to differ from those recorded by another, and different observers can produce different impressions of the same situation. This problem becomes worrying when time comes to use observation as a method for collecting data. This suggests that the data are reliable to be inconsistent between researchers -too dependent upon the individual and the personal circumstances of each researcher. It implies that different observers will produce different data. Therefore, to overcome this intricate problem, researchers have thought to adopt the observation schedule. The use of a schedule aims to minimize, and possibly eliminate, the variations that will arise from data based on individual perceptions of events and situations. More precisely, this instrument targets to provide a framework for observation which all observers will use, and which enable them to do systematically these matters:

- Be alert on the same activities and be looking out for the same things;

- Record data systematically and thoroughly;

- Produce data which are consistent between observers; and with two or more researchers who witness the same event recording the same data.

To achieve these three aims, observation schedules contain a list of items that operate something like a checklist. The researcher who uses an observation schedule will monitor the items contained in the checklist and make a record of them as they occur. The process of systematic observation then becomes a matter of measuring and recording how many times an event occurs, or how long some event continues. In this way, there will be a permanent record of the events, which should be consistent between researchers who use the schedule, because what is being observed is dictated by the items contained in the schedule. When researchers are properly trained and experienced, there should be what is called high 'inter-observer' reliability.

With regard to the advantages of this data collection method, the main merit of observation is that it allows researchers to see directly what people do without having to rely on what they say to do. In such a sense, this method can provide an objective account of events and behaviours that reports provided by second-hand sources do not manage to yield. More interestingly, observation can give descriptive contextual information about the setting of the targeted phenomenon. This can enhance the information a researcher often needs to get more specific data of what he/she is investigating. For other researchers, this way of doing and collecting information can become more reliable and its results valid if structure is added by means of using observation schedules. For some, the use of this specific technique provides answers to the problems associated with the selective perceptions of observers and hence this can lead to produce more objective observations. In precise words, the schedule efficiently eliminates any bias from the current emotions or personal background of the

observer. It gives a means for collecting substantial amounts of data in a relatively short time span. And when properly developed, this technique should achieve a high level of inter-observer reliability in the sense two or more observers using a schedule record very similar information.

Contrary to these positive claims, a number of raised questions on this method remain unanswered. For many researchers, these unanswered points are regarded as drawbacks. For instance, one can namely first refer to the idea that sees that within the sphere of observation what is often targeted is only what is observed, whereas in practice many of the key variables and processes that researchers investigate are mental and thus unobservable. In addition to this, the second problem is that recording a phenomenon does not necessarily lead to understanding the reasons why it has happened, particularly when low inference categories are used. Another more specific and technical third problem concerns the presence of the researcher in the area where the investigation is taking part and where his/her task is to observe some given phenomena. For many experts, the presence of the researcher can, in some way or another, affect and bias the participants' behaviours. That is, the quality of observational data is dependent on the skill which the researcher uses to conduct his/her observation. Finally, a serious additional disadvantage with observation is that regardless of coding convention applied it involves a reduction of the complexity of the observed situation, and by focusing on the target categories the observer can miss some important features. Highly shared schedules also share the weakness of qualitative measures, namely that the examined categories are preconceived and the instrument is not sensitive to context-specific emergent information.

In the present study, because the main purpose of the researcher was to explore the teachers' and students' perceptions, attitudes, behaviours, and reactions in relation to the Current Testing System, and further to implementing the ATM, the researcher acted as a

'non-participant' observer, standing in the back of a classroom and observing how things were occurring in this setting. Choosing this type of observation aimed at eliminating any sort of influence, or negative attitude towards the presence of the researcher on the part of the teachers and students. Moreover, it is worth mentioning that the researcher's previous experience as a secondary school teacher during a period in his teaching career helped him to behave in a natural way in the context of the research and with the participants that gave him the opportunity to gain time and in the mean time collect useful information.

2.1.2.2.4 Focus Groups

A focus group is a research method used to collect data through a group interaction on a topic determined in advance by the researcher. Dörnyei (2007) recognized that focus groups are sometimes treated as a subtype of interviewing because both the format and the interviewer's role considerably resemble to some extent to what is taking part in the interviewing process. Cohen et al. (2005) converge with this view. They add that two groups are a form of group interview, though not in the sense of backward and forward between interviewer and group. Rather, the reliance is on the interaction with the group who discusses the topic supplied by the researcher. Hence, the participants interact with each other rather than with the interviewer, such that the views of the participants can emerge - the participants' rather than the researcher's agenda can predominate. It is from the interaction on the group that the data can be collected.

In a focus group, the format is based on collective experience of group brainstorming, that is, participants thinking together, inspiring and challenging each other, and reacting to the emerging issues and points. This within group interaction can yield high-quality data as it can create a synergistic environment that results in deep and insightful discussion (Dörnyei, 2007: 144). Individual interviews might put a great deal of pressure on the relation between the interviewer and interviewee; on the other hand, a 'focus group' can provide prompts to

talk, correcting or responding to other plausible audience for that talk that is not just the researcher. So, focus groups work best for topics people could talk about to each other in their daily lives - but do not (Chinda, 2009: 116). For Dörnyei (2007), the format allows for various degrees of structure, depending on the extent to which the researcher relies on an interview guide - or as he names it a protocol - rather than giving the participants freedom to discuss some broad topics - just like with one-to-one interviews, the semi-structured type of focus groups is the most common format because it includes both open-ended and close-ended questions posed by the researcher.

In terms of characteristics, a great deal of researchers have agreed upon that in order to make focus groups suit the requirements of an adequate data collection method, the latter needs to specify some basic characteristics. Dörnyei (ibid) summarizes these characteristics in the following points:

- **Size**: The size of a focus group has to range between 6-10 (sometimes 12) people. Fewer than six people would limit the potential of the "collective wisdom", whereas too large size makes it difficult for everyone to participate.

- **Composition**: It has been found that the dynamics of focus groups works better with homogeneous samples. Therefore, in order to obtain a wide range of information, the usual strategy is to have several groups, which, as a whole, are different from each of which is made of similar people; this is usually referred to as 'segmentation' and it involves within-group homogeneity and intergroup heterogeneity in the sample.

- **Number or parallel focus groups**: The standard practice is to run several focus groups in any research project; in this way, we can mitigate any idiosyncratic results that occur because of some unexpected internal or

external factors that affect the dynamics of a group. Thus, in order to achieve adequate breadth and depth of information, it is usually recommended that a project involves 4-5 groups as a minimum, with a few more if possible.

Like previous data collection methods, focus groups present strengths, as well as they underlie weaknesses. About the strengths, it is agreed on that the merits of this method are displayed at two different levels. In educational research, focus groups can be used for programme evaluation to evaluate the effectiveness of a particular course to understand what was or what was not working, and eventually strive to show the reasons of each of the stimulations. In a limited scope of study in applied linguistics research, focus groups have been widely used for generating ideas to inform the developments of other data collection methods such as questionnaires and interviews. For Morgan (1988), focus groups can be useful tools for:

- orientation to a particular field of forms;

- developing themes, topics, schedules for subsequent interviews and/or questionnaires;

- generating hypotheses that derive from the insights and from the group;

- generating and evaluating data from different sub-groups of a population and;

- gathering feedback from previous studies;

Because of the above useful uses of this data collection method, many researchers point that focus groups could be also helpful to triangulate with other more traditional forms of interviewing, questionnaires, observation, and eventually other methods.

On the other side, the main weakness of the focus group is that it is time-consuming to set-up and conduct, and that it requires a good communication skill on the part of the researcher. There is also more need to improvise because the number of questions to be asked following the interview guide is relatively low and a lot of content emerges from well-managed group discussion facilitated by probe questions. Another additional weakness concerns the issue of transcribing in that it is argued that it is usually difficult to transcribe because of the number of people. For this reason, it is advised to use video-recordings of the interviews to overcome this limitation, so that the moderator can identify who is speaking at any one time.

In the present study, as stated above in 2.1.2.2.4, the researcher employed this data collection method in the present study for two main reasons. First, after a first round of the individual interviews, focus groups aimed at encouraging participants to share their ideas and experiences in providing data on the ground of the researcher's set out aims, and in some perspective, this allows the participants to reflect their views on the subject in more comfortable way. Second, because of the flexible and information-rich nature of the method, focus groups were used to triangulate with the other data collection methods used in this study in order to provide more reliable results.

2.2 Research Phases

The present study was carried out in three successive phases. Taking into account the limitations of the study and mainly time constraints, the feedback was held during two school years: 2010/2011 and 2011/2012. In precise terms, the three phases were fulfilled as follows:

Phase One: From January 2011 to June 2012

This phase was designed to collect about the issues of teaching and testing English as a foreign language in the Algerian secondary schools. The main concern of this phase was to

see whether there is any washback   of the Current Testing System on teaching and learning. In this respect, during this particular phase of the research, the present researcher came to:

- collect information on EFL teachers' behaviours in the actual teaching/learning context through classroom observation.

- collect information on the students' behaviours in the natural teaching/learning context through classroom observation.

- collect information on EFL teachers' perceptions to the Current Testing System by which those teachers constantly evaluate and assess their learners' achievement in accordance with the instructional objectives set out in the used syllabus through a questionnaire (01) for the teachers.

- collect information on EFL students' perceptions to the achievement tests their teachers permanently develop during the school year to evaluate and assess their achievement in that foreign language through questionnaire (01) for students.

At the end of Phase One, the researcher collected all the necessary data that helped him to understand how EFL teachers in the context under study teach the subject of English as a second foreign language, and in the meantime, how EFL students within the same context learn that foreign language. More importantly, these collected data gave the researcher a clear picture on how both teachers and students in the Algerian secondary school perceive the Current Testing System, which is used to evaluate and assess the learners' degree of achievement in the taught subject.

All the information and data that were collected in Phase One are meant to be the core of the next chapter (Chapter 3) in this thesis that mainly deals with exploring the context where teaching and learning take place and testing is held to serve its right purpose. The Chapter's Title is "*Teachers' and Students' Perceptions of, and Attitudes towards Teaching and Testing EFL in the Algerian Secondary Schools*".

Phase Two: From September 2011 to June 2012.

This second phase was designed to implement the ATM. To put this model into practice, achievement tests, drawn out from this proposed model, were developed by the researcher with the collaboration of EFL teachers in Biskra Region and then asked a selected number of EFL classes to sit for these tests. These administered tests were proposed to students at the end of three successive trimesters of the school year 2011/2012, in parallel with the achievement tests that were constructed by the EFL teachers, subject of this study, based on the Current Testing System. In precise terms, the students sat for achievement tests from the Current Testing System a week after they were asked again to sit for achievement test proposed by the ATM.

In accordance with the implementing of the ATM, the researcher continued to attend sessions with the selected classes, as a non-participant observer, during the time that followed the trialling of these newly, modified achievement tests. The basic aim of this procedure was to collect data from the real context about the changes in attitudes and behaviours of the teachers and students that might have occurred after the introduction of these new tests.

All the information and data that were collected in Phase Two are meant to be the core of chapter 4 that deals particularly with the issue of implementing the ATM. The chapter's title is *"Towards an Intervention: The Alternative Testing Model"*.

Phase Three: From June 2011 to June 2012.

This third stage overlapped with the second one. It was designed to collect data on the teachers' and students' perceptions of, attitudes and reactions towards the implementation of the ATM. The gist of this phase was to see whether or not there is a washback effect of this new testing system on teaching and learning. In this respect, during this third and last phase in this study, the present researcher came to provide:

- A survey of teachers' attitudes towards the implementation of the ATM.

- A survey of students' attitudes towards the implementation of the ATM.

- A survey of teachers' reactions towards this innovative act.

- A survey of students' reactions towards this innovative act.

To obtain and collect the required information and data and results in this mixed-methods approach, also known as triangulation, the latter consisted of employing four data collection methods. These included the following:

- Administration of Questionnaire (02) for Teachers;

- Administration of Questionnaire (02) for Students;

- Interview for Teachers (02);

- Interview for Students (02);

- Classroom Observation; and

- Focus Groups (also known as Group Discussion) with Students only.

All the information and data that were collected in Phase Three are meant to be the core of Chapter 5, which specifically deals with teachers' and students' reactions to the ATM. The chapter's title is *"Teachers' and Students' Reactions to the Alternative Testing Model Implementation"*.

Below is a summary of the Three Phases in this research and the relevant methods employed to collect data (Figure. 2.2).

I- Phase One:

    Stage: " Teachers' and Students' Perceptions of, and Attitudes towards Teaching and Testing EFL in the Algerian Secondary Schools.

    Time: January 2011- June 2012

    Methods:

        - Questionnaire for teachers (01).

        - Questionnaire for students (01).

        - Classroom observation.

_____

II- Phase Two:

    Stage: "Towards the ATM Implementation"

    Time: September 2011- June 2012

    Methods:

        - Classroom observation.

_____

III- Phase Three:

    Stage: "Teachers' and Students' Reactions to the ATM Implementation"

    Time: September 2011- June 2012

    Methods:

        - Questionnaire for teachers (02).

        - Questionnaire for students (02).

        - Classroom observation.

        - Interview with teachers (02).

        - Interview with students (02).

        - Focus groups with students (02).

Figure 2.2: Research phases and the used data collection methods.

2.3 The Ethical Issues in this Study

Throughout all the stages of this study, the researcher strived to take into account the Ethical Guidelines for conducting the present research. He also ought to be aware of the adoption of explicit conformance to the common Ethical Research Practice. That is, the

researcher has greatly considered the effects of the research on participants and has acted in such a way to preserve their dignity as human beings. For a number of researchers, this is the required ethical behaviour. In a broad sense, ethics has been defined as:

> A matter of principled sensitivity to the rights of others. Being ethical limits the choices we can make unlike truth is good, respect for human dignity is better, even if, in the extreme case, the respect of the human nature leaves one ignorant of human cited. (Cavan, 1977)

The present research conformed to the general ethical guidelines that accounted for ethical issues, such as, privacy, anonymity, and confidentiality. Cohen et al. (2005 : 62) define these latter and provide a clear picture on how and what a researcher should consider. First, privacy is usually thought of as a behaviour which extends to all information relating to present physical and mental conditions, personal circumstances, and social relationships which is not already in the public domain. Second, anonymity is seen as the right to protect the research participants and the relevant data they elicit from being personalized and identified. The essence of this is that information provided by participants should in no way reveal their identity. Third, confidentiality is another way of preserving a participant's privacy. This simply means that although the researcher knows who has provided information given, he will not make the connection known publicly.

On the ground of the above ethical issues, the present researcher abided by an ethic of respect of participants, knowledge, and quality of research throughout the investigation. This means that all the data, findings and conclusions are reported in a clear way and with a great detail to avoid any sort of ambiguity or confusion, and even misunderstanding or misuse, and allow others either readers or researchers to understand and interpret the data. The researcher strived to attach all the required documents and applications forms in the appendix in order to theoretically ground the employed methods and justify the choices and interpretations. Moreover, the researcher insisted on to be a transparent and accurate as possible by

incorporating all relevant details on matters, such as the context, respondents, instruments, and methods used to carry out the investigation and obtain the required results.

About the research participants, all of them were informed about any step that was carried on in the research. To make things clear enough for anyone of the participants from the very beginning, the researcher sought to obtain their consent in advance. That is, before starting any of the pre-set tasks, the research participants were informed about the research to be undertaken, including who the researcher is, why the study is being done, the length of time required to fulfil the tasks. However, it is worthwhile that proceeding in such a way did not directly imply that the researcher revealed the intentions and purposes that are expected to be attained. This was for the sake to avoid any sort of influence or bias that can generate from the researcher's way of doing things. Furthermore, the researcher was quite aware that all the participants who took part in the research were not forced to participate. The participants were told in advance that they could quit whenever they thought they diverged with the process of the research. In this way, the researcher sought to make the participants feel at ease and comfortable, and what they are supposed to do is certainly what they are doing at their will. In accordance with the guidelines, the researcher submitted consent letters for participants to sign. In other cases, the participants were asked only to give their oral consent. What is significant on this last point is that no matter what the form of the consent that was undertaken i.e.; oral or written, the researcher decided that the universal standard norms of ethics had to be fully respected and obtained in all cases to make the research data collection methods and their relevant research methodology credible. A copy of the consent letter used in this study is shown in Appendix 1.

In practical concerns, in the present research, following the ethical guidelines illustrated and discussed above, the experts, who evaluated the contents, formats, and the degree of congruence of the questionnaires to the research purpose and questions, the students and

110

teachers, who completed the questionnaires either in the piloting stage or in the final study, did not give their names when time came to collect these questionnaires. A copy of the questionnaire validation form is displayed in Appendix 2. The main purpose of this procedure was not to identify the participants as it is recommended in the ethical guidelines. With the other data collection methods, such as classroom observation, interviews and focus groups, the researcher opted for pseudonyms to refer to the participants. For example, teachers were given the pseudonyms as T_1, T_2, T_3, etc…, and students, S_1, S_2, S_3, etc. Moreover, because he carried out this study in a context where some subsequent aspects, such as gender and cultural background needed to be treated with some care, the researcher objected to use some techniques, mainly video films during classroom observations and with individual interviews and focus groups in order to avoid any polemic or tension among the participants. To compensate this shortage, the researcher only relied on audio recordings with the hope that they could provide the desired outcomes. Above all, all that was required on the part of the researcher is that the study conformed to a large extent to the general guidelines of ethics.

2.4 Epistemological Positioning and Roles of the Researcher

It is crucial to discuss in the last section of this chapter the roles of the researcher because they directly affect the data as well as their analysis. First of all, it is worth mentioning that before graduating at the university level, the researcher worked as a secondary school teacher in one of the secondary schools in Biskra Region for about ten years. This teaching experience in this context played a major role and contributed greatly in eliciting the problematic on which the present study rests and also was an important source of information that helped and guided in the ongoing of the investigation. During the investigation, the researcher was recognized as a colleague by the teachers who took part in the different stages of the study. This intimate relationship between the researcher and the

participants in the study made it easier for the researcher to cooperate and collaborate for the sake of overcoming the myriads of obstacles and problems that he encountered on the ground. Besides, because of his former status, the researcher was allowed to attend the weekly coordinating meeting that EFL teachers held in their schools. The researcher's presence in the teachers' room was seen by teachers of other subjects to be quite ordinary, and hence did not pose any problems since no one of them complained about it. In short, the researcher in this context was accepted as member of the secondary school community.

Moreover, because of the researcher's previous experience as a secondary school English teacher, it was easy for the researcher to understand and even predict some behaviours of both teachers and students. However, this does not mean that the researcher acted and behaved in a way that might have influenced or affected the participants to make the data biased. Rather, he insisted on a crucial point that the people who are involved in this study should act and behave naturally as if the researcher were not present. All this was done to ensure a certain high degree of the researcher's credibility and objectivity.

Conclusion

This chapter summarized the theoretical and methodological considerations for this study. It indicated that a mixed-methods approach was appropriate for the investigation, since each single method has its individual weaknesses. In addition, this study is a longitudinal qualitative research that integrated also quantitative, ethnography, and case study strategies. The quantitative study was adopted to elaborate and validate the results obtained by qualitative instruments. This strategy aims to overcome any weakness that can come up from the reliance only on one approach. Besides, this study is ethnographic as the researcher was in the research settings as an observer to explore the context where the investigation took part and hence enable himself to observe the participants' perceptions of, attitudes towards, and

behaviours of, to the prevailing Current Testing System and afterwards their reactions to the implementation of the ATM. At another level, the adoption of a case study aimed at yielding a thick description of the situation of the participants. To collect data, the researcher used questionnaires, observation, interviews, and focus groups as data collection methods.

The following chapter is about the Preliminary Study, and the results the researcher is intended to obtain after putting the data collection methods into practice.

Chapter Three: Teachers' and Students' Perceptions of, and Attitudes towards Teaching EFL and the Current Testing System.

Introduction

This chapter is devoted to report a Preliminary Study conducted to display in some detail the participants' perceptions of, and attitudes towards both teaching and learning English in the Algerian secondary schools and the Current Testing System adopted by EFL teachers in this context to evaluate and assess their learners. The beginning of this chapter is to elicit the fundamental research questions and purposes that guided the investigation in this initial phase of the study. The second section attempts to describe thoroughly the research design that the researcher employed to obtain insightful data. Next, a third section is wholly about the results of this Preliminary Phase from the two data collection methods: questionnaire, and observations that were utilised on the ground to gather information. Fourth, a following section discusses the findings of the Preliminary Study and draws a clear picture on the phenomenon and the context under study. Finally, the last section summarizes all what has been arrived at in this chapter.

3.1 Research Questions and Purposes in Phase One

The literature in the field of exam washback suggests that one of the first steps in the study of exam washback is to conduct a preliminary or exploratory study (Wesdrop 1982; Tsagari 2006). For a great deal of researchers, the preliminary studies help to resolve uncertainty regarding whether the existing testing systems would exert any sort of washback that could yield the situations that are certainly prevailing in the context under study. Besides, this particular initial phase would also give a transparent picture on the matter of how testing is frequently carried out and the specific purposes it targets. In short, this part of a study is usually used to pave the path to a better understanding of the context where both teaching and

learning take part; and, in the meantime, how the participants in this operation perceive the way by and through which they are constantly evaluated and assessed in accordance with the relevant instructional objectives.

The Preliminary Study reported here had the practical aims of gathering evidence in order to:

- identify the aspects of teachers' behaviours in the classroom that needed to be studied in relation to the Current Testing System. These included:

  - teachers' strategies in designing and developing achievement tests,

  - teachers' learning strategies, and

  - teachers' teaching strategies.

- displays students' perceptions about aspects of classroom teaching in relation to the Current Testing System. These included:

  - students' perceptions on achievement tests by which they are constantly assessed and evaluated,

  - students' perceptions about their teachers' teaching methodology, teaching activities, and the use of English inside and outside the class, and;

  - students' perceptions of their motivation and opinions about instruction in that subject.

Furthermore, in the Preliminary Study, the researcher would be able to report practical insights that confirm or disconfirm the existence of any washback effect of the "Bac" exam on the teaching and learning purposes and their relevant achievement tests that EFL teachers in the Algerian secondary schools use to evaluate and assess their students. In case, this hypothesized assumption is confirmed, there would be a need to make clear enough the following matters:

- establish the existence of the washback effect of the "Bac" exam on teaching, learning and testing,

- arrive at a clear understanding of its nature,

- provide the opportunity to deepen the research plan by sharpening ideas, and

- suggest directions about what needed to be investigated further.

As a matter of fact, the Preliminary Study in the present research addressed the following research questions:

- Does washback from the "Bac" exam exist?

  If it does, what is its nature and scope on:

  - the contents of teaching (eg. curriculum, teaching materials)?

  - the methods of teaching?

  - classroom assessment?

  - student learning?

  - feelings and attitudes of teachers and students?

The questions posed previously were derived from the researcher's first attempt at conceptualizing the research problem. They were mainly based on the insights gained when reviewing the relevant literature review (Chapter 1), but were also due to current practices of EFL teachers in the Algerian secondary schools observed by the present researcher during his teaching career at this level for 10 years.

The process of formulating the above research questions was in line with Watanabe (2004) and Tsagari (2006) who advise that in case where tests constructors' intentions of their tests "… are absent, not clearly defined or not clearly articulated", as it might be in the Preliminary Study, then, a description of washback and the context where teaching and learning take part is very crucial in order to better understand the nature of this phenomenon for further exploration.

116

3.2 Research Design of the Preliminary Study

3.2.1 Data Collection Instruments

3.2.1.1 The Questionnaires

3.2.1.1.1 Rationale and Aim

Teacher's Questionnaire (01) (TQ1)

The purpose of the teacher's questionnaire (TQ1) is to explore the context where the teaching/learning process of EFL in the Algerian secondary schools is carried out. Also, it investigates how EFL teachers in this context proceed to design and develop one type of language tests, achievement tests, in order to evaluate their students' attainment in English language in accordance with the contents of the syllabus and their subsequent instructional objectives. In sum, the teacher's questionnaire (TQ1) intends to check out whether or not EFL teachers in the Algerian secondary schools follow systematic procedures to develop achievement tests. More importantly, the adoption of this data collection method in the Preliminary Study seeks to see whether or not any washback effect- of the Current Testing System has occurred on the teachers' perceptions of classroom teaching and testing.

Student's Questionnaire (01) (SQ1)

The purpose of the student's questionnaire (SQ1) is to investigate the students' perceptions and behavioural differences that could be linked with classroom teaching and learning as they are practised by EFL teachers. It also tackles the subject of achievement test by which EFL students in this context are constantly evaluated and assessed with. Meanwhile, it looks for how this testing instrument is congruent with the teaching programme and its relevant instructional objectives.

3.2.1.1.2 Structure and Content

Teacher's Questionnaire (01) (QT1)

The teacher's questionnaire (TQ1) in the Preliminary Study is consisted of 3 parts and was designed in English.

Section One (TQ1-17) consisted of 17 questions and 52 items. In this questionnaire, two types of questions were presented. The first type was about close-ended questions, with a space for detailed justifications, and the second type was open-ended questions. Close-ended questions are often welcomed because they can help to gain data on given topics in order to reduce the problems of respondents guessing answers. On the other side, the open-ended questions are seen to be attractive because they invite an honest, personal comment from the respondents. This part of the questionnaire deals with aspects related to the issue of test design. Crucially, the test design stage is where one needs to describe in detail the components that will correspond to the degree of achievement students are making in that foreign language. The product of the design stage includes the following matters:

- A description of the purpose (s) of the test.

- A description of the characteristics of the test takers.

- A description of the constructs to be measured.

- Considering some qualities of test usefulness.

Section Two (TQ18-33) consisted of 16 questions and comprised 64 items. Like the previous part, this second part is made-up of questions of two types: close-ended and open-ended questions. The gist of this part deals with aspects related to the test development stage, also known in other areas as the operationalization and administration stages. The test development stage is where one needs to display whether or not teachers consider the crucial components and steps that are thought to be an integral part of the operationalization and administration processes. The product of this particular stage includes the following points:

- A description of the format.

- A description of the test specifications

- A description of the testing methods.

- A description of the major skills.

- A description of the scoring method.

Section Three (TQ34-36) consisted of 03 questions and comprised 03 items. The questions in this part are designed in an open-ended format. In this part, the teachers were asked to provide their own opinions on aspects related to the degree of congruence between the achievement tests they frequently design and develop and the contents of the taught programme.

What is notable in the preliminary teacher's questionnaire Part 4 (TQ37-40) is that it consisted of 4 questions and the same number of items. In precise terms, this part was developed in order to enable teachers to express their opinions on the contents and lay-out of the questionnaire. It stood as an opinionaire. The final outcome of this procedure was to propose some useful recommendations for the researcher in order to help him remedy the pinpointed anomalies, overlapping, and useless ideas within the different items throughout the questionnaire. What is worth mentioning is that, in the final version of the questionnaire, this fourth part was completely supressed. In the final version of the teacher's questionnaire (TQ1) for the Preliminary Study, 3 themes were elicited. Each one of these themes has a strong link with one of the research questions stated in section 3.1 All these themes are summarized and illustrated in Table 3.1.

Structure	Contents
Theme One	**Test Design (TQ1-17)** • A description of the purposes of the test. • A description of the characteristics of test takers. • A description of the contents to be measured. • Considering some qualities of test usefulness.
Theme Two	**Test Development (TQ18-33)** • A description of the test format. • A description of test specifications. • A description of the testing methods. • A description of the major skills. • A description of the scoring method.
Theme Three	**Teachers' Perceptions of the Current Testing System (TQ34-36)** • Perceiving how is the degree of congruence between test contents and the contents of the taught programme. • Perceiving the degree of validity and reliability of the developed tests.

Table 3.1: The Preliminary Study structure and themes of the teacher's questionnaire (TQ1)

Student's Questionnaire (01) (SQ1)

The student's questionnaire (SQ1) for the Preliminary Study consisted of 2 parts and was designed in English. However, it is worth mentioning that this questionnaire was translated into Arabic, the native language of the questioned students, and handed to these students in this language. The rationale behind translating the original version into the students' native language was to minimize misunderstanding, reliability, and validity problems caused by the language factor (Geisinger, 1994). In addition, by doing this, the

researcher aimed at making the respondents generate their ideas at maximum without being frustrated from the handicap that the non-mastery of that foreign language may pose. In terms of questionnaire structure, the student's questionnaire (SQ1) consisted of 2 parts with 19 questions and 58 items (SQ1-19). Section One (SQ1-16) consisted of 16 questions and comprised 55 items. The main purpose of this part was to collect useful information on the study of attitudes towards the achievement tests their teachers usually develop for them to judge their achievement in accordance with, and its relevant instructional objectives. The type of the questions in this first part in the questionnaire was of a multiple-choice questions format.

Section Two (SQ17-19) is consisted of 3 questions and the same number of items. In this second part of the questionnaire, the students were expected to provide their own views on matters related to the Current Testing System. In some details, they were invited to say whether or not this adopted and practised testing system is useful. In case their answers were negative, they were asked to suggest some remedies that they think could make the system become more effective and useful. The questions in this part were designed in an open-ended format.

Like the teacher questionnaire (TQ1), in this first phase of the study, the third part of the student's questionnaire was designed with the purpose to judge the contents and lay-out of this data collection method. As carried out previously with (TQ1), the last part in the preliminary student's questionnaire stood as an opinionaire. For that reason, 4 questions were suggested to students. All these questions were designed in an open-ended format. What is notable with this part is that it was omitted in the final version. Furthermore, in the student's questionnaire (SQ1) for the Preliminary Study, 2 themes were targeted. In their gist, these specific themes have a direct link with the research questions stated in section 3.1 those themes are summarized and illustrated in Table 3.2.

Structure	Contents
Theme One	**Students' Attitudes towards Achievement Tests (TQ1-16)** • Collect information on how students view achievement tests they are constantly assessed and evaluated with.
Theme Two	**Students' Perceptions on the Current Testing System (TQ17-19)** • Collect information on how students perceive the Current Testing System.

Table 3.2: The Preliminary Study structure and themes of the student's questionnaire (SQ1)

3.2.1.1.3 Piloting and Validation of the Instrument

To make this data collection method ready for final application, the researcher opted to piloting the two questionnaires (TQ1) and (SQ1) for the Preliminary Study. The piloting study was carried out in June 2011. The basic aim of this crucial stage in questionnaire development was to ensure validity and reduce problems with the procedure of administering the questionnaires. The participants in this stage were experts from different universities in Algeria and another one from a university in the U.K. The rationale behind submitting the adopted data collection method to local and foreign experts for expertise was to maximize their validity in order to obtain comprehensive outcomes. Besides, the two cited questionnaires were also handed to a number of EFL teachers and students in the Algerian secondary schools from the population of the study in Biskra region. What is notable about this last point was that the asked teachers and students were not chosen from the sample of the study. This, of course, is what is recommended in research methodology and by research methodologists. Again, this was fulfilled in such a procedure in order to follow meticulously

the systematic steps that should be undergone in a pilot study. In sum, all the followed stages in the process of developing the two questionnaires for both teachers and students aimed at making this data collection method valid, and in the meantime, help the researcher gain insightful and useful data about the subject under exploration.

In practical concerns, the teacher's questionnaire (TQ1) was submitted to a foreign university expert. He holds a PhD degree in English language studies from the London Metropolitan University. Similarly, the same questionnaire, i.e. (TQ1), was submitted to 2 Algerian experts. Like the one from the U.K., these 2 experts hold a PhD degree. The first was from the University of Sidi-bel-Abbes (Algeria), and the second from the University of Adrar (Algeria). Besides, the teacher's questionnaire was also given to 10 EFL teachers from the secondary schools in Biskra region. On the whole, 13 people participated in expertizing and reviewing the teacher's questionnaire (TQ1) in the Preliminary Study. Table 3.3 below displays the participants in this piloting stage of the teacher's questionnaire.

Nº	Participant	degree	Place of work
01	E_1	PhD	- London Metropolitan University.
02	E_2	PhD	- University of Sidi-bel-Abbes.
03	E_3	PhD	- University of Adrar.
04	10 EFL Teachers	"Licence" degree	- Secondary Schools in Biskra region.

Table 3.3: Participants in the pilot Study for the teacher's questionnaire (TQ1)

In terms of remarks and comments of the experts concerning the contents and format of the teacher's questionnaire (TQ1), they were as follows:

E_1 thought that there is a need to reformulate some questions. He saw that each of the proposed question needs to be clearly linked to one of the research questions. He added that the questions that do not bring new insights should be dropped out. The same expert assumed

123

that when some items do not pertain in a way or another to what the researcher is searching, in this case, there is a need to leave them.

In addition, he raised an important remark on the questions of the questionnaire. He noted that most of these questions seem to be about the issue of test design and development. For him, it sounded more likely since in this questionnaire, the researcher is looking at the effectiveness of the test, there is more responsibility to overlook questions and items about the teachers' view on general pedagogical approaches used by the teachers. Due to this significant remark, 9 questions in Section Two (TQ8-16) in the preliminary teacher's questionnaire were completely omitted.

E_2 proposed that it is better to reduce the number of questions, for he believed that when a questionnaire sounds long enough for the respondents, this would lessen the degree of their attention and concentration. He added that the more a questionnaire is lengthy, the more it looks boring to answer.

E_3 gave an insightful remark in that he noticed some questions that do not pertain to the research purposes that the researcher ought to answer. He proposed that it is more appropriate to avoid them. The latter can bring more consistency to the questionnaire. As a response to this raised remark, the researcher suppressed Q8 - Q9 - Q11 - Q12 and Q13. This remark converged with the one suggested by E_1 so far.

Contrary to the wealthy comments and even criticisms yielded by the university experts, the reviewing of the teacher's questionnaire (TQ1) by secondary school teachers, subject of this pilot study, did not bring new ideas since the majority of these teachers completed the questionnaire and did not give any significant comments. A final copy of teacher's questionnaire (TQ1) in the Preliminary Study is shown in Appendix 3.

At another level, about the preliminary study student's questionnaire (SQ1), the pilot study was carried out at the same period of time as the teacher's questionnaire (TQ1). The

participants in this stage were as follows: 1 General Inspector from the Ministry of Education, 2 University Experts - 1 from the University of Oum-El-Bouaghi, and another from the University of Adrar - and 10 EFL teachers from the secondary schools in Biskra region, and the same number of EFL students from the population subject to this research. In sum, 23 people participated in expertizing and reviewing the Preliminary Study student's questionnaire (SQ1). Table 3.4 summarizes the participants in the pilot study in this first phase.

Nº	Participants	Degree	Place of work
01	E_4	General Inspector of English	Batna
02	E_5	PhD	University of Oum-El-Bouaghi
03	E_6	PhD	University of Adrar
04	10 EFL teachers	'Licence' degree	Secondary schools in Biskra region
05	10 EFL students	Students	Secondary schools in Biskra region

Table 3.4: Participants in the pilot study for the student's questionnaire (SQ1)

Referring back to the participants' comments and remarks, those varied from one participant to another. On the whole, they were as follows: the General Inspector of English thought that the student's questionnaire (SQ1) seems to be too long. For him, it would be economic to fuse some questions and reword the ones that are considered to bear some ambiguity. Surprisingly, the remarks raised by E_4 in their details were general and he did not clearly point out which questions had to be fused and which ones should be reworded. For E_5, she expressed some doubts about the terminology present in the student's questionnaire; she advanced that to answer the items, the questioned students might take too much time to fulfil this task. She estimated this time to be more than half an hour for each one of the respondents. In this respect, for E_5 spending more than half an hour filling and answering the questions could create some boredom and anxiety that often lead the respondents to leave the

questions unanswered. If he completely agrees on with these doubts raised by E_5, the researcher thinks that this problem had been settled down so far since the submitted questionnaire to students was translated into their native language. In doing so, the issues of language difficulty and its influence on the time spent by the respondents to answer the questions are no longer posed.

In the same sphere, E_6 thought that some items are overlapping; those latter bear quite the same ideas to be made more explicit by the respondents. As a remedy to this problem, E_6 suggested, for instance, that Q13 and Q31 could be melted in one question. Likewise, the same procedure should be carried out with items Q15 and Q32. Besides, E_6 added that it would be more appropriate if Q17 and Q19 could be reworded to bear a more straightforward sense. This, for sure, leads the respondents to avoid any sort of confusion that is often the result of ambiguity in the meaning of the questions due to formulation. As what has been done previously with the EFL teachers about their comments on the teacher's questionnaire (TQ1), the teachers did not make significant remarks that could bring some changes to either the form or contents of this data collection method. They merely filled in the questions, and in some cases, they left some questions unanswered without putting any comments to justify their act. Contrary to the EFL teachers, EFL students who participated in this pilot study shared the common idea that the questionnaire was loo long because it comprised a great deal of questions. They proposed it would be better if the researcher proceeded to reduce the number of questions and make the questionnaire shorter. In addition, 2 students from the 10 participants indicated that some of the questions needed to be reworded to facilitate access to their sense. However, on this point, these students did not point out which of these questions they targeted. As a result to the comments and remarks raised by the participants who took part in this pilot study, the researcher proceeded to some adjustments and modifications to some of the questions and sections in the questionnaire for the student's (SQ1).

As an example, Q31 and Q32 were reworded and Parts One and Two in the preliminary student's questionnaire were completely reviewed. Also, some minor changes were made in relation to the way the questions were asked; this was done to make the structure of the questions more formal. The final version of the student's questionnaire in the Preliminary Study is displayed in Appendix 4.

3.2.1.2 The Observation

3.2.1.2.1 Rationale and Aim

As mentioned in section 2.1.2.2.2, there is a well-documented literature on the different approaches to classroom observation. In particular, two basic approaches have been usually considered the most available, structured and unstructured approaches. A structured approach will know in advance what is looked for and will have its observation categories worked out in advanced. However, an unstructured observation will be far less clear on what is looked for and will therefore pave to go into a situation and observe what is taking place before deciding on its significance for the research (Cohen et al., 2005: 305). The weaknesses of these two approaches have been debated (Allwright & Bailey, 1991, Denscombe, 2010); Structured observation records what happens, but not the intentions that motivated the behaviour. Besides, unless a researcher is about exactly to observe and design a well-tested observation scheme, the subtleties of the situation can easily be ignored. With detailed field notes in an unstructured observation, researchers may be less sure about what exactly the focus of the observation is at the beginning and try to clarify what behaviours to focus on as they gain experience during the observations. In this way, the data collected from an unstructured observation usually relies on the researcher's inferences and detailed field notes in a particular context.

3.2.1.2.2 Structure and Content

What is worth noting is that to carry out an observation investigation, it is crucial to design an observation scheme. For research methodologists, a research scheme is indispensable to collect systematically insightful information about the matter under study. However, even though a great deal of observation schemes have been developed based on the research and findings in the domain of language teaching, a few of these schemes can be appropriately used in relation to the area of language testing. Thus, in order to overcome this shortage in observation schemes in the field of the present investigation, the best solution would be to adapt the available instruments to suit what is intended by and expected to this study to realize in its terminal stage. Therefore, on the ground of this claim, the classroom observation scheme thought to be the most appropriate and adopted by the researcher in the present research is the one that was borrowed from a well-known model, the Communicative Orientation of Language Teaching, also referred to as the COLT, which was produced by Spada and Fröhlic (1995).

As discussed earlier in Chapter 2, in terms of structure, the COLT scheme is divided into two parts, often labelled A and B. In Part A, classroom events are described at the level of an activity. Each activity is timed so that a calculation of the percentage of time spent on the COLT features can be determined. For the analysis of this part, the use of check marks of those features describing the activity makes it possible to get an overall picture of each event in the classroom. In part B, in the COLT scheme, there is more focus on verbal exchange between teacher and students or among students themselves as they occur within each activity. For the analysis of this part, Part B is analysed according to each activity. Check marks are counted and divided with verbal amount of marks in the same group.

In the present study, Part A of the COLT was employed since a classroom analysis at this level sought to describe whether there is communication between teacher and students,

and students with students. In the meantime, classroom analysis also intended to find out whether classroom activity matched the nature of the research questions to be answered. Part B was not adopted because simply the nature of this investigation is not on the issue of the language to be used. (See Table 3.5 below).

Time	Participant Organization		Activity Type	Activity Content	Material Used					
	Whole Group	Individual			Type			Purpose		
	T to S/C	S to S/C			W	A	V	P	S	N

Table 3.5: Classroom observation scheme in this Study

T to S/C: Teacher to Students or Class as a whole   **V**: Visual
S to S/C: Students to Students or Class as a whole   **P**: Pedagogical
W: Written   **S**: Semi-pedagogical
A: Audio   **N**: Non-pedagogical

In terms of pure description of this research's observation scheme, this scheme consisted in the following five categories: time, participant organization, activity type, and material used. To make this observation scheme more practical and easy for use, the researcher proceeded to code these categories. Such a procedure enables to gain much space and time while operating the observation task. The researcher had only to tick under the category of participant organization and materials used during the observations. He also strived to make detailed field notes under the category of time, type of activity, and activity content. By this process, the researcher aimed at making the observation operation very systematic and theoretically grounded. These are crucial conditions in any scientific investigation.

Next, what is below is some detailing of the categories that make up the observation scheme used throughout this study. Hopefully, this additional information about the

observation scheme in this research can give more insights and understanding about the scheme's categories and their functioning:

1. **Time**: This category concerns mainly analysing the observed lessons in terms of time spent during an activity. To fulfil this initial step, the analysis is based on the notion of a segment. In this content, a 'segment' is defined by Mitchell, Parkinson, and Johnston (1981: 12-14) as "a stretch of classroom discourse having a particular topic and involving participants (in the case of the present research, these are teachers and students) in carrying out an activity or task through interaction". A change of topic/activity type or mode of interaction indicates a completion of the start of new segment. (Gibbons, 2006:05).

2. **Participant Organization**: This second category deals basically with the pattern of organization for classroom interaction (Allen et al., 1984:235). The three patterns are:

   - Whole-Class: involving teacher to students, or students to students.
   - Pair or Group Work.
   - Individual Work.

   The investigation of this category can give a clear picture on how teaching and learning are taking place in a classroom. Observing the participant organization enables the researcher to investigate the interaction patterns in classrooms; and, of course, this can help him to provide a comparative evaluation of the interaction of the participants before and after implementing the ATM. In the Preliminary Study, this particular category can also be useful to draw a transparent picture on both teaching and learning and how they are being performed. To make this last point clear, at the level of this category, the researcher has to

answer the following questions: Is teaching communicative? And does interaction occur because of communication?

3. **Activity Type**: This category checks out what teaching and learning activities as a percentage of class time are. The aim is to look more closely at the type of activities that are currently performed in the classroom. Besides, the other aim is to seek to describe whether each activity among the set of the used activities can provide insights on how teaching and learning are carried out.

4. **Content**: This category investigates the contents of activities. The study of the contents may help the researcher to evaluate the degree of relevancy between what is being taught and how this latter is assessed.

5. **Materials Used**: The last category in this series of categories in the research's observation scheme sheds light on the types of teaching materials that are used in teaching and learning and the purpose or purposes those particular types intend to achieve. In this sense, these include the following:

(1). Types of materials include the following:

- Written materials (textbooks, workbooks, dictionaries …).

- Audio-materials (songs, extracts of talks …).

- Visual materials (PowerPoint, movies, documentaries; …).

(2). Purpose of materials concern the following:

- Pedagogical materials (textbooks).

- Semi-pedagogical materials (exam papers).

- Non-pedagogical materials (material intended for non-school purposes).

131

3.2.1.2.3 Piloting and Validation of the Instrument

It is worth noting that classroom observation was the data collection method that was carried out by the researcher throughout the whole period of the fieldwork. In other words, the researcher observed the selected classes from a period of time that began early in the 2010-2011 school year until the end of the next school year in 2011-2012. Purposefully, this long period of observation aimed to collect useful information from the ground before and after the time of the ATM implementation in order to depict any changes in the participants' behaviours and attitudes towards the teaching and learning of EFL in the context under exploration. Certainly, the task of observing classes was not an easy matter because of some reasons. Some of these were that, for instance, the present researcher needed to obtain a permission to get access to the different secondary schools that form the sample of this study. Obtaining permission required long administrative procedures that began with the Educational Department in Biskra and arriving to the headmasters in the targeted secondary schools. Another procedure that needed to be completed is to get consent from the teachers and students to attend sessions. The researcher managed to obtain both oral and written consent, but with the condition that he was not allowed to video-film full or part of the attended sessions. This objection was manifested by the fact that the common school regulations forbid such type of practices. To compensate this shortage, the researcher relied only on the observation scheme and the process of observing.

Concerning the issue of piloting and validation of the observation instrument, this latter was undertaken through systematic stages. First, the scheduling of each classroom session was agreed upon in advance between the researcher and teacher. All the school visits were pre-arranged with the teachers involved according, of course, to these teachers' timetables. Reminder telephone calls were frequently made one day before school visits. Second, 2 secondary school teachers, who were not part of the final sample in this research, volunteered

to co-attend the observation sessions with the researcher during this pilot stage; they knew all the process of observing and how to fill in the observation scheme. The advantage of their presence was to compare the field notes they took with the ones taken by the principal researcher. This aimed at reaching a high degree of agreement of the data collected via observation, and hence realize more credibility of the final outcomes of this data collection method. At the end of the piloting stage of this classroom observation, the suggested observation scheme for this study was kept as it was without any modifications. The pilot stage confirmed that the observation scheme based on Part A of the COLT scheme is the mechanism by which the researcher is expected to use for the collection of information.

3.2.2 Respondents

The population in the Preliminary Study consisted of EFL teachers and students from random sample of schools and classes in Biskra region. To get access to these schools and to be in touch with the targeted population during this phase, a request was submitted to the Education Administrative Authority in Biskra, wherein a road map comprising the main purpose of the school visits, the targeted EFL classes and population were identified. After a short period of time, a positive response to this request was given. Thanks to the Educational Administration in Biskra, that the researcher got comprehensive statistics in which useful details about the number of schools, teachers, and EFL classes were displayed.

On the ground of this platform, the researcher worked to select the population with opportunity and convenience. He sought to recruit the participants in this initial phase on the basis of different background variables. The most crucial of these were that EFL teachers should be from the two sexes and having different teaching experiences. In the same vein, EFL classes should be selected from different secondary schools in Biskra region. The procedure also aimed to create a balance in the schools' geographical distribution. Finally, a random sampling of the students was used to ensure that various respondents of the whole

population were represented. The respondents included students with different psychological and social backgrounds. Below, some of the main characteristics of these participants are shown and presented in relation to the data collection method employed.

3.2.2.1 Respondents to the Questionnaires

There was a total of 20 teachers. They were all surveyed using a questionnaire. All of them completed the questionnaires and handed them back on time, representing a response rate of 100%. At this first phase of the research, the participants were chosen from the whole population of teachers who teach English language in the secondary schools in Biskra region. This means that the selected sample included teachers who teach EFL for secondary classes and teachers who teach English for EFL classes. In doing so, the researcher aimed at extending his sample to other teachers to collect an amount of information and see how teachers of English in the Algerian Secondary schools design and develop achievement tests. The same strategy was also used with students. In sum, the number of the questioned students was exactly 100. All of them completed the questionnaires and handed them back on time.

3.2.2.2 Respondents to the Observation

Because the researcher attended observation sessions with the targeted teachers and classes as a non-participant observer, the researcher selected to observe 8 classes from 6 secondary schools in Biskra region. Before attending the observation sessions, he managed to arrange the sessions according to the teachers' timetables. Above all, in the preliminary phase, the researcher attended 68 sessions and observed 8 teachers and about 300 students in their classes learning English language.

3.2.3 Data Collection Procedures

3.2.3.1 Data Collection Procedures for the Questionnaires

To submit the questionnaire for both teachers and students, many procedures were followed. The first and most common procedure was to hand the questionnaire to the teachers who themselves handed them to their students. Another procedure was to send the questionnaire for some teachers via email. They were asked to fill it and email it back using the same medium. The third and final procedure consisted in distributing the questionnaire to the teachers during an in-training service that was organized by a General Inspector of English language. This last procedure was possible with the help of the local educational authorities and the General Inspector of English.

3.2.3.2 Data Collection Procedure for the Observation

Before attending the observation sessions, the researcher informed the teachers that his presence was only to sit in the back of the classroom and observe what is happening in the classroom and take notes. He added that while observing, he had not the right to intervene or make comments. Besides, he told the teachers that they had the right to look and check the notes he had taken. The teachers were informed that they would be provided with a summary of the results of his study. The aim was to ensure confidentiality and conform this research to the ethical guidelines.

3.3 Results of the Preliminary Study

3.3.1 The Questionnaires

3.3.1.1 Teachers Questionnaire (TQ1)

Item 1:
What do these words mean to you?

> a. **Test.**
>
> b. **Test design.**
>
> c. **Test development.**

As to this item, teachers provided different answers. The first term 'test' was defined as 'examination', 'evaluation' and 'assessment'. About the other terms 'test design' and 'test development', the questioned teachers gave different answers. Some defined the two terms as 'test conception', and others referred to 'test construction'. For a few the respondents 'test design' and 'test development' are about the process of 'test planning'.

Item 2: Do tests provide you with appropriate information?

Nº of item	Number of teachers	Percentage
A	12	60%
B	04	20%
C	03	15%
D	08	40%
E	09	45%
F	01	05%

Table 3.6: Information provided by tests.

60% of the teachers we questioned agreed to say that achievement tests provided them with information about the students' level because it is the focal point in constructing a test for whatever objective. All the other statements are by-products. 20% of the respondents advocated that such tests gave them insightful information about the teaching programme since it is the only source through which they could be informed about how much achievement their students are making in accordance to the programme taught. Only 15% of the teachers we questioned said that achievement tests could enable them to know whether or

136

not the used syllabus is appropriate or not. They argued that a good test is not only the one that assesses their students, but also it can transgress to evaluate the syllabus on which instruction rests. 40% of the respondents believed this type of tests could provide them with insightful information about the effectiveness of the instructional objectives since one basic requirement of achievement tests is that they must be related to the objectives of instruction. 45% of the teachers that we questioned agreed to say that no one of these variables could be discarded. Rather, they thought that all of them should be considered in order to get a transparent picture and with appropriate information about these tests. For them, each variable is present in order to complete the other ones.

Item 3:
When developing achievement tests, what parameter do you take into account?

As to this item, what could be retained is that 75% of the teachers we questioned left the question unanswered. The very few respondents claimed that the parameters that often took into consideration when designing and developing achievement tests are the ones included in item 2. In other words, this refers to the students' level, the teaching programme, the content of the syllabus, and the instructional objectives. However, no one of them gave detailed, complete, and thorough explanations on how these parameters came to be taken into account in test design and development.

Item 4: When developing achievement tests, do you consider your students general language ability?

Nº of item	Number of teachers	Percentage
A	15	75%
B	05	25%

Table 3.7: Students' general language ability.

75% of the respondents answered positively to this question. They merely asserted that a test developer, the teacher in the case of this research, had to take this feature into account since it would be non-sense if the students' general language ability were fully or partly

eschewed when describing achievement tests. They argued that to consider this criterion obviously stimulates students and make them more motivated to undertake the test. This could be explained by the fact that the students would get the feeling that they were tested in what is supposed that they should be tested on. They also thought that considering the students' general language ability when designing achievement tests make the latter accessible and not beyond those students' level. At the end, those tests would gain more validity and consistency of the final scores.

Item 5:
What abilities do your tests target?

This item was originally meant to tell whether or not the teachers we questioned define as clearly as possible the construct to be measured, keeping in our minds that this last component is crucial in the test design stage. About the teachers' answers, the respondents agreed on that the basic abilities their tests target were mainly the four skills i.e., listening, speaking, reading, and writing. They all noted that the tests they developed focused on other language components such as grammar and lexis. However, the majority of teachers we questioned made clear for us that in practice many of these skills and abilities are completely discarded. They explained that, for example, they never tested their students on the two skills, listening and speaking. In this way, they believed that their tests were not as congruent to what they presented in instruction.

Item 6:
Before designing achievement tests, what theoretical bases do you rely on?

Surprisingly, none of the respondents gave an answer to this question. All the teachers we questioned left this item unanswered.

Item 7: Do you link you test-items in achievement tests to the content of the syllabus?

N° of item	Number of teachers	Percentage
A	19	95%
B	01	05%

Table 3.8: Degree of relevance between test-items and syllabus.

138

The majority of teachers we questioned answered positively this item. They argued, in the case of achievement tests, teachers are compelled to link the test-items to the content of the syllabus. They added that it is unfair and not right to assess the students on something they have not learnt in class.

Item 8:
Are your tests quantity-based or quality-based as far as the syllabus is concerned?

Some teachers claimed that their tests should be quality-based without completely neglecting the amount of knowledge learnt. However, they added that it is not quite possible to include everything dealt with in class in the test. They argued that in order to make things sound more practical, it is better to proceed to a systematic sampling of the most important and frequent language components to be incorporated in the test. Other respondents believed it is more plausible to create an equilibrium between on one hand, the quantity of the knowledge learnt, and on the other hand, the quality of what has been dealt with in class. For them, being selective is the key to make tests look valid.

Item 9: Do you find scores often relevant to the students' achievement and progress in accordance with the instructional objectives?

Nº of item	Number of teachers	Percentage
A	05	25%
B	10	50%
C	02	10%

Table 3.9: Degree of congruence between students' scores and achievement in English.

From the above results, we can see that 50% of the teachers said that the scores did not usually reflect the actual achievement the students were making. Nevertheless, the great majority of the respondents did not provide explicit and clear justifications to this phenomenon. The only reason they gave was that most of the students lacked a considerable basic knowledge in English. A fact which often hampered them to do their best in learning that foreign language.

Item 10:

When developing achievement tests, do you consider the relationship between test-content and the instructional objectives?

As to this item, the majority of the teachers we questioned answered positively.

However, they did not provide any explanations about how they conceived this relationship.

Item 11: When developing achievement tests, do you take in account your students' …?

Nº of item	Number of teachers	Percentage
A	02	10%
B	14	70%
C	09	45%
D	02	10%
E	05	25%
F	03	15%

Table 3.10: Students' characteristics.

From the above results, only 10% of the teachers we questioned said that they took into consideration the personality characteristics of their students. However, no one of them explained thoroughly how s/he usually came to consider this intention. The majority of the respondents 70% thought that among all these variables it is background knowledge which is the crucial element in the design of achievement tests since it strongly contributes in making the test effective in evaluating the degree of the students' achievement. 45% of the teachers we questioned claimed the dire need to consider the feature related to the aptitude for learning because it is that element which often urges any given student to learn English language. 10% of the respondents believed that taking into account the native language in the test design is important. They argued that the majority of students in the Algerian secondary schools usually refer to the mother tongue to learn a foreign language. 25% of the teachers we questioned thought a good and reliable test is the one that does not discard any variable because as teachers, we should consider the learners as human beings and everything contributes to their success. What is noticeable on this item is that the majority of the

respondents did not give thorough details about their answers. Instead, they merely stated that one or all the variables are crucial in test design.

Item 12: Do you think you achievement tests can serve your teaching?

Nº of item	Number of teachers	Percentage
a	14	70%
b	04	20%
c	02	10%

Table 3.11: Test impact on instruction.

All the teachers we questioned seemed to agree to say that tests could serve their teaching. 70% of the respondents asserted that tests often exerted an influence on their teaching in that a test is not an end in itself. Rather, it is a source through which a teacher could get feedback that needs to be capitalized on. For them, a test should be a continuation to instruction, not a buffer zone. Besides, the majority of the respondents believe that such tests could provide them with appropriate means to remedy their teaching. They allow a better achievement and progress of students.

Item 13: Do you take into consideration the affective side of your students when designing and developing achievement tests?

Nº of item	Number of teachers	Percentage
A	02	10%
B	18	90%
C	00	00%

Table 3.12: Students' affective factors in test development

All the teachers we questioned were unanimous (90%) to say that they did not take into account the affective side when designing and developing achievement tests. They argued that these negative answers were justified by the fact that giving too much consideration to affection would make this type of tests lose their validity and therefore this reality leads to the provision of non-credible final scores. On the other side, few respondents 10% thought that this aspect had to be taken into consideration because students are the ones around whom all the factors turn. They are the targets of tests. Their positive attitudes vis-à-vis tests help them achieve better scores and make substantial progress.

Item 14: Do you think your students learn from achievement tests?

Nº of item	Number of teachers	Percentage
A	08	40%
B	12	60%

Table 3.13: Test impact on students.

Most of the teachers we questioned (60%) claimed that it is usually not evident that their students learn from tests. They believed so because their students thought that all what is included in the test are only relevant to the test. Others believed that even if their students localized and recognized their mistakes through tests, they still continued doing the same faults. Nevertheless, 40% of the respondents declared that achievement tests could be a very useful means to learn. For them, tests could be one way to review and consolidate previously learnt knowledge.

Item 15: Do you think achievement tests concern other broader areas?

Nº of item	Number of teachers	Percentage
A	02	10%
B	18	90%

Table 3.14: Test impact on other broader areas.

90% of the respondents thought that achievement tests could not concern other broader areas such as education system, society and test users. Rather, they believe those tests are only limited to classroom uses. They argue that the only parties who are concerned by those tests are students and their teachers. 10% of the teachers we questioned asserted that those tests could transgress to affect other parties and areas. They advocated that those tests could provide test developers with useful information about ways to improve the format and content of tests in order to enable learning to be more effective. In addition, if students get good results, this way may provoke a change in their attitudes towards school. Test users will also benefit from the outcomes of those tests to evaluate the relevance, effectiveness and validity of the education system as a whole.

Item 16: Do you think you need to be assisted in order to develop achievement tests?

Nº of item	Number of teachers	Percentage
A	07	35%
B	08	40%

Table 3.15: External assistance in test development.

35% of the teachers we questioned thought they needed to be assisted when designing and developing achievement tests. They believed that a teacher, at any stage of his/her career, needs assistance of some sort. They advocated that nobody could claim absolute expertise in the matter. Such an assistance could be made through a positive cooperation with the other teachers. 40% of the respondents thought differently. They believed that there was no need for any external assistance because this type of tests did not require a huge amount of resources either material or human. They added that any classroom teacher could prepare and carry out these tests alone.

Item 17:
What kinds of resources are not available to carry out your tests?

Concerning this item, the majority of the teachers we questioned agreed to say that the main resources that are not available to carry out their tests are particularly those that are needed in testing the two skills: listening and speaking. Those latter include authentic materials, audio-visual and laboratories.

Item 18: Do you produce texts for you tests?

Nº of item	Yes	No	Sometimes
A	10%	80%	10%
B	10%	70%	20%
C	05%	90%	05%
D	40%	40%	20%

Table 3.16: Test format.

What comes out from the results embodied in the above table is that 80% of the teachers we questioned did not use their own productions, did not provide a list of glossary to explain the different words in the test-text, and did not care a lot about the style of the language in the test they presented to their students the day of the test.

143

Item 19: What is / are the source (s) of your test-texts?

Nº of item	Number of teachers	Percentage
A	12	60%
B	04	20%
C	02	10%
D	02	10%

Table 3.17: Text Sources.

From the above table, one can see that (60%) of the teachers we questioned used the textbooks on which instruction was based to create or extract texts for tests. 20% of the respondents referred to other external sources such as newspapers and magazines to choose a correspondent text for the test. 10% of the teachers used new technological sources like internet to select texts. A few number of the questioned teachers thought it would be better to rely on personal productions.

Item 20: What parameter (s) do you consider in choosing your test-texts?

Nº of item	Number of teachers	Percentage
A	12	60%
B	15	75%
C	11	55%
D	09	45%

Table 3.18: Text selection parameters.

From the gathered results, one can see that 60% of the respondents took into consideration the content of the programme when choosing texts. This made the latter more accessible to students. Besides, it allowed the evaluation of the achievement students were making. 75% select texts taking into account the students' level. This choice might decrease the degree of frustration and anxiety students often meet the day of the test. 55% of the respondents considered that the length of texts should correspond to the time allowed. 45% of the respondents agreed to say the fact that all these variables were crucial in choosing any text. They added that the teacher had to create an equilibrium among the parameters to yield a final text attractive and useful for the purposes of the test.

Item 21: Do you specify the vocabulary included in the test-text from the content of the syllabus?

Nº of item	Number of teachers	Percentage
A	04	20%
B	14	70%
C	02	10%

Table 3.19: Vocabulary specifications.

70% of the teachers we questioned did not specify the vocabulary included in the text from the content of the syllabus. They did not proceed to any vocabulary-specifications because they thought it is difficult to meet exactly the kind of texts that conformed to all the criteria put forward. 20% of the respondents believed that there had to be vocabulary specifications before developing tests. They claimed that those latter should be at least genuine. They might contain already known words as well as some new ones. Consequently, if students were tested on a certain amount of learnt vocabulary, this would urge them to pay more attention to everything they saw in class. 10% of the teachers we questioned thought that the specification of the vocabulary to be included in the text from the content of the syllabus was not necessary.

Item 22: Do you test lexis …?

Nº of item	Number of teachers	Percentage
A	19	95%
B	01	05%
C	00	00%

Table 3.20: Vocabulary testing methods.

95% of the teachers we questioned tested lexis in relation to the reading passage because students were expected to provide their comprehension of the text and their understanding of the lexical meaning of items from context. These teachers asserted that deducing the meaning of the lexical items from the context allowed the development of the students' mental ability to learn and acquire knowledge. 5% of the respondents test lexis independently from the text because one of the aims of achievement tests is to evaluate the amount of vocabulary memorized by students.

Item 23: Do you test grammatical structures on the basis of …?

Nº of item	Number of teachers	Percentage
A	16	80%
B	04	20%

Table 3.21: Amount of structures in texts

The above table shows clearly that 80% of the teachers we questioned said that they tested grammatical structures on the basis of the structures taught in class. They argued that this procedure usually enabled them to test what had been taught. Besides, it allowed them to have more opportunities to select the structure or structures they thought they needed to test. 20% of the respondents asserted that a testing structure included in the test-text prevented students from anxiety and frustration.

Item 24: Do you limit the number of grammatical structures to be tested?

Nº of item	Number of teachers	Percentage
A	20	100%
B	00	00%
C	00	00%

Table 3.22: Number of structures in tests.

All the teachers we questioned answered positively. They justified their answers by the fact that one could not test everything in one test. All parts of a test should be balanced. Besides, they asserted that the time allotted to the test prevented the testing of a large number of structures.

Item 25: When testing the grammatical structures, what do you consider?

Nº of item	Number of teachers	Percentage
A	08	40%
B	07	35%
C	05	25%

Table 3.23: Testing structures frequency.

40% of the questioned teachers believed it is crucial to keep the same structures in each test because this allowed better achievement. However, they insisted on the assumption that it is better to vary the testing techniques in order not to enable students to remain fixed on the same structure; some others (35%) thought that varying the structures from one test to

another is the only way to evaluate accurately whether the students had grasped what they

have learnt or not.

Item 26: When testing grammatical structures, do you use the same testing technique (s)?

Nº of item	Number of teachers	Percentage
A	14	70%
B	04	20%
C	02	10%

Table 3.24: Testing grammar methods.

From the above results, we can see that 70% of the teachers we questioned believed

that they always used the same testing methods when testing grammatical structures. They

asserted that teachers were bound to train structures to be familiar with these testing

techniques since they are the one that are permanently proposed in the "Bac" exam. 20% of

the respondents claimed that they rarely used the same testing techniques. They thought it is

quite an obligation to vary the methods since the structures they teach are various. For them,

the variety of structures allowed the avoidance of boredom on the part of students. 10% of the

teachers used the same techniques in order to provide a thorough evaluation.

Item 27: Do you choose the testing techniques because they are …?

Nº of item	Number of teachers	Percentage
A	06	30%
B	05	25%
C	15	75%
D	02	10%
E	02	10%

Table 3.25: Testing grammar selection procedures.

Out of the above results, one can see that the majority of the teachers chose the testing

techniques that are part of the ones that are usually proposed in the 'Bac' exam manual. They

added that their primary focus is to train their students to answer the typology of tasks that

currently the 'Bac' exam papers include. Their argument was that if students got bad scores

on this national exam, it is the teacher who is going to be blamed.

Item 28: Do you test the listening skill?

Nº of item	Number of teachers	Percentage
A	00	00%
B	20	100%
C	00	00%

Table 3.26: Testing listening.

All the teachers answered negatively to this question. All of them agreed to say they never tested their students on the listening skill. The main argument they advocated was that their concern was to prepare their students to answer the typology of tasks and activities that are currently included in the 'Bac' exam papers and since those latter do not incorporate the testing of listening, the questioned teachers thought there is no need to care about this matter. For some other questioned teachers, the reason that prevented them to develop a listening test is mainly due to the unavailability of the required human and logistic resources that are compulsory to make this test true. However, it is worth mentioning that the totality of the teachers who were questioned expressed their desire to see one day that their tests include a part wholly devoted to testing the listening skill because if realized then, for sure, this makes the tests they frequently develop look more useful and credible.

Item 29: Do you test the speaking skill?

Nº of item	Number of teachers	Percentage
A	00	00%
B	20	100%

Table 3.27: Testing speaking.

All the questioned teachers answers negatively. They thought that there is no need to test this skill because their students are not tested on it in the final 'Bac' examination. Although most of them believed that this is not a good way of assessing the actual level of their students.

Item 30: Which skill do you think is possible to test within you context of teaching?

Nº of item	Number of teachers	Percentage
A	11	55%
B	05	25%
C	03	15%
D	03	15%

Table 3.28: Teachers' perceptions of testing listening and speaking.

From the above table, we can see that 55% of the questioned teachers thought it is possible to test the listening skill in the context of their teaching. 25% of the respondents believed they could do the same thing with speaking. Those teachers claimed it is an easier matter to incorporate a listening test in the achievement tests they usually developed; and, in the meantime, it is more complicated to do the same way with the speaking skill. Their argument on this last assumption was that the common crowded classes stood as a handicap towards the implementation of a speaking test. However, almost all of these respondents viewed that it is unfair to teach something without testing it.

Item 31: Do you proceed with the same scoring method?

Nº of item	Number of teachers	Percentage
A	07	35%
B	06	30%
C	07	35%

Table 3.29: The employed scoring methods.

Concerning this item, 35% of the respondents stated that they usually followed the same scoring method when developing achievement tests. The aim was usually to train their students to respond to the typology of tasks and activities that are currently included in the 'Bac' exam papers. 30% of the teachers we questioned did not proceed with the same scoring method. For them, the latter often rested on the importance assigned to each part in the test. For instance, if the test focused more on grammar, it was crucial to score this language component higher. For the last alternative answer, 35% of the teachers linked the scoring

method to various criteria such as the difficulty, weighing, and importance of each test-item in the test.

Item 32: Do you inform your students about the criteria of your scoring method?

Nº of item	Number of teachers	Percentage
A	02	10%
B	12	60%
C	03	15%

Table 3.30: Students' perceptions of the scoring method: An evaluation.

60% of the questioned teachers said that they did not inform their students about the scoring method. They assumed that if their students are frequently informed about how they were scored, this made these students focus more on specific test-items that had higher scores and ignore those that had lower scores. However, other teachers (10%) maintained that students could trust their scoring if they were told in advance about it. A small number of the teachers we questioned (15%) said that they sometimes informed their students about the scoring procedure. However, in the meantime, they did not provide any additional explanations or details about their answers.

Item 33:
How does this affect your students?

Our respondents gave different answers to explain how informing the students about the scoring method procedure affects them. For some teachers, this fact is positive because their students would get the feeling that their teachers shared with them everything. This might create a strong bridge of trust and confidence between the two as partners. For others, they bore a different view in that they thought that their students, if, usually told about the scoring method used in the test they design, they would not have the same focus towards the different test-items. This, in itself might narrow the areas learnt in lessons.

Item 34:

In your opinion, do you think the teachers in the secondary schools give importance to testing?

Most of the teachers we questioned said they did not actually give the right importance to this matter. They asserted that they had never been trained in how they should develop valid and reliable tests. For them the issue of testing is merely limited to assigning scores to their students; and, at the end, it is a question whether to reward or sanction their students.

Item 35:

Do you think the Current Testing System applied on the Algerian secondary schools should be modified, kept, or completely change it with another one more useful?

The majority of the teachers who were questioned thought it is better to modify this Current Testing System.

Item 36:

What would you suggest as far as the testing system used in the Algerian secondary schools level is concerned?

To answer this item, myriads of responses were proposed. We shall report some of them below in this list:

- "… the Current Testing System should take into account the listening and speaking skills."
- "… the Current Testing System should encourage the students to monitor their own achievement."
- "… time allotted for the tests should be reviewed."
- "… why not suggest some tasks based on translation."
- "… the serious revision of the Current Testing System must provide an appealing format, a consistent, and a fair scoring for all the parts in the test."
- "… testing should be regarded as a paramount partner to teaching."

3.3.1.2 Students Questionnaire (SQ1)

Item 1: Do achievement tests cover the content of what you have learnt?

N° of item	Number of students	Percentage
A	21	21%
B	19	19%
C	51	51%
D	01	01%

Table 3.31: Correspondence between test-contents and syllabus

Among the 100 respondents, 51% of the students we questioned thought that achievement tests, by which they were usually assessed, covered a little part of the programme they have dealt with in lessons. Only 21% of those students saw that there was considerable correspondence between the test-content and the syllabus. 19% of the questioned students considered that this correspondence is somehow acceptable. However, 7% of the respondents did not see any correspondence.

Item 2: are the texts of these tests based on the content of the units you studied in class?

Nº of item	Number of students	Percentage
A	06	06%
B	39	39%
C	55	55%

Table 3.32: Correspondence between test-contents and lessons.

39% of the students who answered this item-question said that the supporting texts included in achievement tests did not correspond to the contents of the taught units/themes. 55% of the questioned students noted that it was not always evident that the contents of the texts of the tests were taken from the contents of the lessons they got in class. However, only 6% of the students we questioned thought there was a good correspondence.

Item 3: Do you think the texts are …?

Nº of item	Number of students	Percentage
A	54	54%
B	45	45%
c	01	01%

Table 5.33: text length.

What we can notice from the above table is that 54% of the respondents thought that the texts of achievement tests, by which they were assessed, were too long. 45% of the students from the same sample reported that the texts were not very long. However, 1% of the students we questioned thought that the texts were too short.

Item 4: Is the language of these texts …?

Nº of item	Number of students	Percentage
A	59	59%
B	40	40%
C	00	00%
D	01	01%

Table 3.34: Text difficulty.

The above results show that 59 students among the 100 respondents thought that the language of texts is very difficult. Their argument was that the language of texts of achievement tests they usually encountered differed from the language they were taught with. However, 40% of the students we questioned provided opposite answers. They believed that they were, to some extent, familiar with the degree of difficulty of the texts of tests. For them, this particular matter did not pose any problems to them.

Item 5: Is the number of difficult words that these texts comprise …?

Nº of item	Number of students	Percentage
A	64	64%
B	30	30%
C	06	06%

Table 3.35: Number of difficult words in a text.

As to this item, 64% of students we questioned thought that the number of difficult words included in the texts of tests was very big. Those respondents saw that these texts of tests often comprised lexical-items that seemed very new to them i.e., they had not dealt with them in class. About 40% of the answers reported that the number was not very big, and therefore, it did not impact on the answers. Contrary to these two comments, 6% of the students we questioned saw that the texts of tests did not include a big number of difficult words.

Item 6: Do you think the content of these texts corresponds you interests?

Nº of item	Number of students	Percentage
A	08	08%
B	70	70%
C	22	22%

Table 3.36: Correspondence between text content and student's interest.

70% of the students we questioned thought that the contents of texts in achievement tests, usually presented to them, did not correspond to their interests. 22% of the respondents from the same sample viewed that the contents of the texts of tests sometimes corresponded to their interests. However, only 8% of the respondents saw that there was a correlation between the contents of texts of tests and their interests.

Item 7: Before you answer the questions, how many times do you read the text?

Nº of item	Number of students	Percentage
A	07	07%
B	11	11%
C	07	07%
D	75	75%

Table 3.37: Number of readings.

The majority of our respondents said that the number of times they read the texts of tests before answering the questions depended primarily on the types of the text. They explained that the more the text was drawn from the contents of lessons, the less number was. 11% of the respondents thought it was more suitable to read twice the text before proceeding to answering the questions. However, 7% is shared by those who read one time and the other ones who read three times the text before going to respond the different item-questions.

Item 8: Does your reading (s) of a text take …?

Nº of item	Number of students	Percentage
A	21	21%
B	10	10%
C	69	69%

Table 3.38: Spent time in reading.

This item was asked to confirm the results we got in item (07). The answers we got in for this item were absolutely in prefect concordance with those ones in item (07). 69% of the respondents thought that the number of readings depended on the nature of the text. 21% of the answers posited that the reading of texts took a long time. 10% of the students thought it was quite the opposite in that their reading took a short time.

154

Item 9: Are the questions often …?

Nº of item	Number of students	Percentage
A	07	07%
B	43	43%
C	50	50%

Table 3.39: Instructions clarity.

The students' answers to this question could be grouped into two main folders: the first one comprises fifty (50) students who claimed that questions were not usually clear. For them the problem resided mainly in the ambiguity of the instructions. In order to respond to the typology of these tasks, they only guessed. The other group of respondents comprises around (43) students. These considered that it was not difficult for them to answer the questions because they believed that the questions were clear enough. They explained that it was thanks to their teachers who usually trained them to answer these tasks. Only 7% of the students who were questioned assumed that the questions were very clear and hence they did not encounter problems to answer them.

Item 10: Are the testing methods used in these tests usually …?

Nº of item	Number of students	Percentage
A	73	73%
B	21	21%
C	06	06%

Table 3.40: Testing methods variety.

73% of the respondents noted that the testing methods their teachers use to assess them were the same in everything. 21% thought that these testing methods varied slightly from one test to another. However, 06% of the students who were questioned asserted that the testing techniques completely varied from one test to another.

Item 11: Is the distribution of test-items among the test parts …?

Nº of item	Number of students	Percentage
A	11	11%
B	73	73%
C	16	16%

Table 3.41: Test-items distribution.

The great majority of the students we questioned (73%) said that the distribution of the test-items per test-sections was not equal. They made it clear enough that they often divided their tests into sections. Currently, the number of sections is three: 'Reading Comprehension', 'Mastery of Language' and, 'Written Expression'. They divided the test-items per sections. They added that commonly the 'Reading Comprehension' section comprises the biggest number of test-items. Whereas, the other two sections included less numbers than the first section. 16% of the respondents saw that it was more appropriate to vary the number of test-items from one test to another. However, only 11% thought that it was better to leave an equal distribution of test-items per test.

Item 12: Which test items are difficult to answer …?

N° of item	Number of students	Percentage
A	60	60%
B	00	00%
C	31	31%
D	10	10%
E	00	00%

Table 3.42: Test-items difficulty.

From the results embodied in the table above, we can say that the majority of the students we questioned (60%) thought that the test-items which included the 'Reading Comprehension' were difficult to answer. 31% of the respondents reported that it was the items testing lexis which were difficult. However, surprisingly, no one of the students we questioned thought that the testing of grammar was difficult. 10% of the students who completed this questionnaire believed that the 'Written Expression' section was difficult.

Item 13: Do you think these tests test …?

N° of item	Number of students	Percentage
A	68	68%
B	05	05%
C	05	05%
D	01	01%
E	21	21%

Table 3.43: Kind of tested ability.

68% of the students we questioned said that they thought achievement tests, by which they were usually assessed, tested the degree of their comprehension. They advocated that the biggest part of test-items these tests comprised were the ones that were typically devoted to the 'Reading Comprehension' section. 21% of the respondents saw that what achievement test tested were the language components, including aspects of language such as comprehension of the reading text, lexis, the mastery of the grammatical structures, language usage, and other components; the rest of the students thought that most of the tests focused on one component at the expense of the other ones.

Item 14: To have good results in these tests, you need to …?

Nº of item	Number of students	Percentage
A	08	08%
B	73	73%
C	19	19%

Table 3.44: Revision strategies.

The majority of the respondents (73%) said that they did not revise continually before taking the test. Rather, they revised only a short time before the day of the test. They argued that experience had shown them that there was no need to carry out these learning strategies since they thought that the contents of tests did not often reflect what they learnt in class. They claimed that those tests tended to judge their general knowledge and proficiency in that foreign language, and the amount of progress and achievement they made in the language. 19% of the respondents did not revise at all. They interpreted this negative behaviour in that they viewed the typology of tasks suggested in tests needed only to guess and not to go back to what they dealt with in instruction. However, 8% of the respondents assumed that revision is compulsory to obtain good scores since without revision students could not understand and answer the questions.

Item 15: Is the scoring method of each test known to you?

N° of item	Number of students	Percentage
A	02	02%
B	61	61%
C	37	37%

Table 3.45: Scoring methods clarity.

61% of the respondents ignored how the scoring methods in tests were developed. They explained that most of the scoring methods varied from one test to another. 37% of the questioned students added that they were not always informed about how they were scored. However, a small number of the respondents (2%) said that their teachers usually told them about how they scored the test-items in the different sections of the test.

Item 16: Does your mark in English usually reflect your progress and achievement in that foreign language?

N° of item	Number of students	Percentage
A	24	24%
B	51	51%
C	25	25%

Table 3.46: Correlation between scores and student's level.

51% of the respondents thought that the scores they obtained in achievement tests did not actually reflect the degree they were making in language. 25% of the students we questioned believed that there was a kind of correlation between the scores and the students' level. However, 24% of the questioned students saw that there was a strong correlation between what they obtained as scores and their performance on tests.

Item 17:

In your opinion, what is negative of these tests?

All the questioned students seemed to agree to say that the negative side in the current achievement tests resided in their contents. They asserted that what they were tested in was not what they studied in class. In a word, they strongly believed that the contents of the tests did not reflect the teaching programme. They also added that they believed that tests were not

developed to assess their progress and achievement in English, but they were devised only to measure the amount of what they knew in English language.

Item 18:

In your opinion, what are some suggestions you can give to your teacher to make these tests more suitable?

The majority of the students we questioned thought that the current achievement tests had to be modified. They claimed that teachers had to develop tests with clear instructions, various activities, a balanced distribution of questions and a well-defined scoring method.

Item 19:

Do you have any additional comment on the subject of testing?

Most of the students said that it is crucial to be trained on how to answer the typology of tasks suggested in the "Bac" exam manual. However, in the meantime, it is also important to be tested through tests that make fair judgment about the progress and achievement they are making in that language.

3.3.2 Classroom Observation Outcomes

The following section reports the findings of four teachers among the eight teachers observed during the initial phase in the present research. The gist of this part is to describe classroom activities in order to identify whether or not the lessons were student-centred or teacher-centred, how learning opportunities were provided, and what instructional materials teachers used in instruction. In particular, the major purpose of this investigation is to describe communication in these classrooms and establish the students' communicative behaviours.

As noted earlier in chapter 2, the method that has been used for this purpose is an observation scheme referred to as the COLT. This scheme has been adjusted for this study, as

its original intention did not include some aspects related to the description of how teachers' language is used (see sub-section 2.1.2.2.3). Thus, the questions asked for each column were:

1. Time: How time splits into different activities of class time?

2. Participants organization: Who is holding the floor during the segment of the lessons as a percentage of class time?

3. Activity type: What teaching and learning is realized through various activities as a percentage of class time?

4. Content: What are the teachers and students talking, reading, and writing, or listening to?

5. Material used: What types of teaching materials are used for what purpose?

3.3.2.1 Participant Organization

Table 3.47 summarizes the interaction patterns of the lessons with T_1, T_2, T_3, and T_4 teaching EFL classes. In this sub-section, two component aspects were observed: (a) is the teacher working with the whole class? And (b) are students divided into groups or are they engaged in individual work?

These particular aspects are presented in Table 3.47 below as:

- T to Student/Class (teacher to student or class as a whole).
- S to Student/Class (student to students or class as a whole).
- Group Work (students are working on a certain task in groups).
- Individual (individual work performed by students).

Number of Teachers	Percentage of Lesson Time				
	Whole Class	Whole Class	Group	Individual	
	T to S/C	S to S/C	00%	10%	100%
T1	90%	00%			
T2	95%	00%	00%	05%	100%
T3	45%	00%	40%	15%	100%
T4	85%	00%	05%	10%	100%

Table 3.47: Participant organization on lessons.

   T to **S/C**: Teacher to Student or Class as a whole.
   S to **S/C**: Students to Students or Class as a whole.
 * Percentages are averaged per person and rounded off to be the nearest number.

The findings from the observed classes made it clear for us that the majority of class time was dominated by teacher-centred approach. For instance, with T_1 and T_2, more than 90% of the time spent by the teachers was devoted to explaining the contents of lessons directly without giving time to students to interact. The same remark could be raised with the other teachers T_3 and T_4. The rationale behind these high percentages is that teachers talk more than students do and students themselves remain passive receivers since they only listen without participating in lessons. In case they attempted to participate, all they do is to answer some of the questions their teachers ask. What is particular about their responses was that most of the time, they were short answers, made-up of fragments of sentences, or sometimes, they were limited to "Yes/No" answers. This might result in students not being particularly motivated to engage in a classroom communication.

However, unlike the previous three teachers (T_1, T_2, and T_4), T_3 spent more time, estimated to 40%, to group work. It seemed that this teacher aimed to bring students to communicate in the target language. This teacher showed a great motivation to bring her students to practise the target language. She seemed aware that one of her major objectives in teaching this foreign language is to urge her students to become competent in communicating effectively with the studied language. In this respect, for her it sounded that group work could be one of the best techniques that might help the students to overcome the difficulties they often encounter to express themselves fluently and accurately in English. Contrary to the students in the other observed classes, students who were taught by T_3 attempted to construct correct sentences and explain their ideas in English. With this class, (15%) of individual work was realized. Although this percentage seemed low, but in comparison to what has been observed in the other classes, it could be said that it is the best of what was observed.

To summarize the findings in the first part in the COLT scheme, what is noticeable was that neither student talk nor group work were observed with almost all the observed teachers

161

and classes they taught. It is certain that teachers dominated most of classroom interaction and that even though these teachers were expected to follow and use the general guidelines and principles that the adopted teaching approach in the Algerian educational system, CBA, underlines, teacher talk was the predominant teaching strategy in most of all the lessons.

3.3.2.2 Activity Types and Content as Percentage of Class

In the present sub-section, the researcher explored two basic aspects observed during this initial phase in this study: activity types and content of class time. As noted previously, to obtain information from this part of the research, the researcher answered two important questions considered to be relevant to the two mentioned aspects. These two particular questions were (1): What teaching/learning is realized through various activities as a percentage of class time? and, (2): What are they listening to? The purpose of looking at activity type in classroom teaching was to investigate what kinds of teaching and learning were held through the different tasks. Similarly, by exploring the content of the activities performed through the proposed teaching tasks in the classroom, the present researcher sought to get a transparent picture on the typology of activities that the observed teachers adopted to enable their students learn English language.

In order to realize the objective of this observation section, the current researcher grouped the different teaching tasks into two parts. One part was about the teacher activities and the other one dealt with the students' activities. What is remarkable about this organization was that the findings related to the content were again reported as a percentage of class time. The analysis of the lessons with the observed teachers displayed (a): What types of activities were carried out in the lessons and how these lessons were segmented according to the percentage of time devoted to them by the observed teachers, and (b): Who was holding the floor and in what ways (see Table 3.48).

Activity type	T₁	T₂	T₃	T₄
Teacher Activities:				
Lecturing	30%	25%	15%	30%
Explaining	30%	20%	20%	25%
Directing	20%	20%	00%	15%
Student Activities:				
Listening	00%	00%	00%	00%
Reading	05%	05%	00%	05%
Vocabulary	05%	10%	10%	10%
Grammar	10%	10%	05%	15%
Speaking	00%	00%	00%	00%
Writing	00%	10%	00%	00%
Group Work	00%	00%	40%	00%
Pair Work	00%	00%	05%	00%
Oral Work	00%	00%	05%	00%
Total	100%	100%	100%	100%

* Percentages are rounded off to the nearest percentage.

Table 3.48: Classroom activities of all the observed teachers as a percentage of class time

What could be retained from the results embodied in Table 3.48 is that the teacher was still the predominant focus of the classes. This suggests a fairly traditional teaching method where the teacher is in charge of the events in the classroom. About 80% of the lesson time was devoted to teacher activities in T_1's lessons. 30% of T_2's activities were spent on lecturing and the same percentage was given to explaining. In terms of the language used by this teacher, what is noticeable was that his language was a mixture of English and Arabic. 20% of the class time was spent for directing. This was in the form of some disciplinary statements from the teacher. In this case, classroom management, therefore, was thought to be a source on input rather than an invitation to communication. About students' activities, a lot of time was spent on language where almost all attention is on form. 10% of the time of the students' activities was devoted to learning grammar. Traditional grammatical exercises mix with lessons about the structure of the language. Concerning the time spent by the students to explaining vocabulary, it was estimated to only 05%. The particularity of the activities on learning vocabulary was that they mainly centred on the idea that students had to

enrich their vocabulary stock, so that they would be able to understand texts when reading and why not construct sentences in the target language. However, it seemed that the intention from learning vocabulary excluded the assumption to use the learnt vocabulary to generate communication.

Following the same strategy of classroom observation, the findings from T_2 lessons showed that the same conclusions as the previous teacher, T_1 were attained. About 65% of the lesson time on teacher activities focused on lecturing, explaining, and managing the classroom. These obtained results clearly indicated that most of the time T_2 teaching was devoted to transmit an amount of knowledge without giving any opportunity to students to negotiate or discuss some of this information. Intentionally or unintentionally, the teacher completely dominated the class, leaving no time to his students to interact or communicate using English language. In the percentage of class time of classroom activities presented in Table 3.48, it is clear that the percentages on group work and pair work are estimated to be null. Again, this is a very significant indication of the ongoing of teaching and learning in the context under investigation.

Contrary to the first observed teachers T_1 and T_2, the obtained percentages sounded a little bit different from those of T_3. With this observed class, it seemed that there was some interest on the students' activities that could enable the students to communicate with the target language. For instance, percentages displayed that about 40% of the student activities were devoted to group work. This very acceptable percentage directly implies that T_3 urged her students to practise interactive activities that could lead them to communicate effectively using the learnt language. Also, this obtained result could imply that the teacher's intention might be to limit the amount of time that she used to explain his lessons to enable his students to share some of this time. Of course, this was the only matter that could clearly differ from the general atmosphere that was prevalent with all the other observed classes. For the other

students' activities, it could be retained that there was a slight increase in the class time devoted to pair work and oral practice. The time for both of these two activities was estimated to 10%. Certainly, even though this time is low, but in comparison to the other observed classes, such a remark indicated that there was a serious attempt with T_3 to provide her students with the opportunity to interact and communicate.

With the last observed class, it seemed that there is a return to the same results and conclusions that were obtained previously with T_1 and T_2. For instance, 70% of the lesson time devoted to teacher activities was about lecturing, explaining, and directing. This very high percentage ostensibly indicated that there was a total dominance of the teacher over the set out interactive activities. As opposed to T_3, with this class, there was no room for group work, pair work, or even any tasks that sought to improve the students' oral practice. About these activities, the percentages are estimated to be equivalent to null. In accordance however, usually with the students' activities category, the percentages about learning grammar and vocabulary are estimated to 10% and 15% respectively. Once again, this indicated that there is a return to a teacher-centred approach where much of the teaching and learning activities were in the hands of the teacher. In other words, this means that there is no place for communicative teaching.

Above all, to summarize the findings from Table 3.48, a cross-comparison of the percentage of the time on students activities in the lessons observed showed that there were no substantial differences among, a least, three teachers. What is significant was that what could be retained from T_1, T_2, and T_4 was that their lessons were not interactive and their approach was more a teacher-centred. Most of their lessons turned around teaching some precise language structures and particular vocabulary-items. On another issue, these lessons did not encourage the students to practise the learnt language or provide them with some useful learning strategies that could enable them to use this foreign language appropriately

inside and outside the classroom. Besides, it is notable that although the adopted syllabus rests on the principles of communicative language, the kind of teaching observed during this initial phase clearly showed that there was an imbalance in the teaching of the four skills of language. That is, for instance, listening and speaking were not paid any importance and hence they were completely neglected and not taught. Instead, most of the teaching was centred on teaching the two language components, grammar and vocabulary.

3.3.2.3 Teaching Materials

From the observed lessons, it was noted that the most used teaching materials in classroom with the four observed teachers was the official textbook. This particular material is designed to guide teachers in their teaching, and in the meantime, help students know the contents of each one of the various strategies in a unit. In addition to this pedagogical tool, T_1, T_2, and T_4 used some visuals such as pictures and cards. Apart from these, nothing else was employed by these three teachers. However, contrary to the previous cited teachers; T_3 used a CD for listening accompanied with handouts. Those latter were employed to orient the teacher towards teaching the listening skill. The particularity of the audio-texts was that they comprised a simple language of native speakers. The teacher's argument on the choice of this simple language was that it was very difficult for his students to catch up and understand more complicated forms of language.

In summary, the types of teaching materials for the observed teachers were not substantially different. Most of them were based on the official textbook and the use of handouts for grammar and vocabulary activities. The used materials matched, to a great extent, the adopted teaching methodology that focused more on teacher talk and the practice of grammar and vocabulary tasks that help students to be well trained to respond to the different activities included in the developed achievement tests.

3.4 Discussion of the Results

The results from the used data collection methods revealed many interesting findings in relation to the teachers' behaviours in the actual teaching/learning context, teachers' perceptions of the Current Testing System, and also students' perception of the achievement tests their teachers usually develop in order to evaluate and assess their attainment in English as foreign language. These findings are reported below:

3.4.1 Teachers' Behaviours in the Classroom in Relation to the Current Testing System :

3.4.1.1 Teachers' Teaching Strategies

• The teachers said that they placed a great deal of emphasis on some language components such as grammar and vocabulary. For these teachers, proceeding in such a manner enables their students to be ready for the tests they usually develop to check out their degree of attainment in that foreign language.

• The teachers recognized their teaching is usually evaluated on the ground of the final outcomes their students obtain at the end of a term.

• The teachers claimed that about the four skills, much more attention was placed on reading, and writing. They argued that to answer the 'Bac' exam tasks, students need to be well trained in reading the supported texts and provide written answers.

• The teaching of speaking skill received a little attention because students are never tested on it. In many cases, teachers omitted the part in the teaching unit that is about listening and preferred to devote its time to teaching something else, mainly teaching grammar and vocabulary.

• The other skill, speaking, was completely avoided. Much of the teaching, as noted previously, was teacher-centred. Students merely listen and when it came to communicate, they answered through telegraphic, short answers.

• To compensate this shortage, teachers said that they often taught what they called 'written pronunciation'. The latter consisted in teaching students how they should pronounce the final 'ed' of the past participles of regular verbs, or the final 's' of the plural forms of regular nouns. At another level, they also taught their students to pronounce some words on the basis of syllable division. Contrary to teaching speaking, the teachers believed that teaching 'written pronunciation' directly contributed to the requirements of the test, since, very often, there is a part in the developed tests which is about testing this language component. In this way, this helps in training the students to respond to that kind of questions.

• Although the proposed teaching methodology relied on the principle of the communicative approach, the teachers said that they usually referred to traditional methods to teach some language components. Their argument was that these methods are compatible with the idea underlying the tests.

• Because of the influence of external examination, teachers stressed that their methodology was affected by the 'Bac' exam which made them do things they would not otherwise do. In absence of the influence of the 'Bac' exam, they would employ a communicatively-oriented methodology, focus on needs analysis, and adopt more authentic materials.

3.4.1.2 Teaching/Learning Strategies

• The teachers believe that the main learning strategy that should be taken into account is to lead their students to memorize repeated items, standardize knowledge, and direct learning.

168

• The teachers felt that their students became dependent on them expecting them to not only teach the language but also train them successfully to pass the exam.

• The teachers used some materials to enable their students to learn that foreign language. Most of the time, teachers referred to past papers and   commercial   publications containing activities or supplementary materials focusing on developing skills needed in the 'Bac' exam. In many cases, teachers talked about the adaptation of the teaching materials to secure the objective of being well-skilful in responding the current activities that come in this testing system.

3.4.2 Teachers' Perceptions of the Current Testing System

3.4.2.1 Teachers' Strategies in Developing their Tests

• The teachers recognized that there was an influence of external examination on their classroom tests. This was well noticeable on the structure and contents of their tests that they develop at the end of a term in the school year.

• In precise terms, teachers claimed that their tests often resemble the 'Bac' exam model in that the tests they often develop comprise a supporting text, one section devoted to testing comprehension and lexis, a second section to testing grammar, and a third one to testing writing.

• The teachers said that they never test their students on the two skills listening and speaking, because they are not included in the 'Bac' exam.

• The teachers relied on the textbook materials for their tests. It was believed that the influence of the 'Bac' exam on the achievement tests was also mediated in their teaching.

• Because they thought that their main tasks is to enable their students to obtain good scores in the 'Bac' exam, this prevented them to innovate in their teaching.

169

• Again, because of the influence of the 'Bac' exam on their teaching and how they come to develop their tests, the teachers restricted the use of testing techniques so that their students could acquire the necessary test-taking skills needed to answer successfully the current questions.

• The teachers were unanimous to declare that they ignore completely about the systematic procedures they have to follow to design and develop a useful language test.

3.4.3 Students' Perceptions of their Teachers' Teaching Methodology, Teaching Activities, and the Use of English Inside and Outside the Classroom

3.4.3.1 Students' Attitudes towards Teaching

• The students showed that because of the teaching methodology their teachers employed in classroom, they felt that such communicative activities, which were not related to the needs of the 'Bac' exam, were not welcomed. They did not see any useful purpose behind their use and looked at them with contempt.

• The students claimed that because their teachers over emphasized on teaching them activities that come in the 'Bac' exam, they became unaware of the importance to develop their skills in listening and speaking. All they thought of was desired to provide them with what helped them to get high scores in the final exam.

3.4.3.2 Students' Attitudes towards the Teaching Activities

• The students believed that there was no need to bring activities that did not relate to the achievement tests and the 'Bac' exam.

• In the meantime, the students showed a negative attitude towards the contents of the lessons their teachers often devised to them because they thought that these activities had nothing to do with what they expected from the learning of that foreign language.

170

3.4.3.3 Students' Perceptions of the Use of English Inside and Outside the Classroom

• The students were sure that the method by which they studied English language did not permit them to communicate effectively outside the classroom.

• The students felt that they were extremely dependent on their teachers and the used textbook expecting them to prepare successfully for the exam without paying attention on how to use the language appropriately inside and outside the classroom.

3.4.3.4 Students' Perceptions of their Motivation and Opinions about their English Lessons

• Because success in the final exam was the main objective of learning that foreign language, this created a tense atmosphere where students felt anxious to learn English.

• Also, this atmosphere made students feel bored and demotivated in the classroom.

• The exam-oriented methodology and the kind of teaching materials did not allow students to enjoy the pleasure to learn the language. Rather, they looked at learning as a means of passing the exam.

3.4.3.5 Students' Perceptions of Achievement Tests by which They are Constantly Evaluated and Assessed

• The common attitude of students towards the achievement tests was that they felt afraid of them because the teachers used these tests to make students study.

• A part of the students got an opposite view in that they thought that these tests helped them to become proficient to sit for the 'Bac' exam and hence increase their chance to pass this decisive exam in their studying career.

• However, it is worth mentioning that the majority of students believed that the scores they often obtained did not reflect their actual level in that foreign language.

3.5 Concluding Remarks

The results obtained from the data collection methods established the existence of washback from the 'Bac' exam. The teachers showed that there was a close relationship between the contents of this exam and the teaching and learning that went on in their classes. In some details, the claims of the teachers clearly indicated that their teaching and learning strategies were dominated by the use of activities and materials that fit the requirements of the 'Bac' exam at the end of the secondary education cycle. This fact is interpreted in terms of the neglect of developing the skills that are not tested on the 'Bac' exam, mainly the listening and speaking skills. In this way, such a fact led the teachers to place more value on the skills and activities that were assessed on the 'Bac' exam, and less value on those that were not tested. Teachers claimed that the ordinary syllabus was relegated to a secondary position giving its place to the exam demands. (see Appendix 5 and Appendix 6)

Similarly, the students' perceptions of the teaching methodology their teachers employ to teach them, teaching activities, and materials used, and their use of English inside and outside the classroom were highly influenced by the requirements of the 'Bac' exam. This particular exam was also seen by students to have detrimental effects on their motivation and opinion towards the learning of English language. Above all, this exam made the students uncomfortable and anxious and this in turn led them to hate studies and tests. Creating feeling of anxiety and making them feel bored and demotivated restricted the desire of students to improve in learning this foreign language.

To summarize, the results obtained in this Preliminary Study called into the validity of the achievement tests EFL teachers in Biskra region usually develop to evaluate and assess their students' attainment in English language. However, before reaching any definite

conclusions, it is necessary to ensure the validity of the participants' assumptions, identify clearly other anomalies in the Current Testing System, and more importantly, identify areas of further study that remedy the pinpointed deficiencies. The aim of the next chapter is to synthesize the results of this first phase in this study in the light of the intentions of the researcher to provide the rationale for an ATM that seeks to repair the set of inadequacies in the Current Testing System.

Conclusion

In this chapter, we have collected data about the participants' perceptions of, and attitudes towards the teaching and learning of EFL, and the testing system adopted in the context under investigation in this initial phase of the present study. The first part in this chapter identified the research questions and purposes that limited the scope of the study in this preliminary stage. The second part described clearly the research design that the researcher employed to obtain data. The design and validation procedures, as well as the rationale, aim, structure, and content of all the used data collection instruments were discussed. Part three of this chapter reported the results the researcher reached after putting into practice all the used tools. Finally, the two last parts in this chapter provided a discussion of the results and a summary of the findings.

After having presented the results, a preliminary picture on the participants' perceptions of, and attitudes towards the teaching operation, and testing mechanisms in the Algerian secondary schools was drawn out. This picture ostensibly showed and confirmed that the hypothesis that posits that the washback effect of the 'Bac' exam on the teaching, learning, and testing in the context under investigation exists.

The following chapter presents and describes the ATM, proposed by the present researcher in order to remedy the anomalies pinpointed in the current testing system.

Chapter Four: Towards an Intervention: The Alternative Testing Model Implementation

Introduction

The present chapter begins with a review of the major findings investigated in Phase One of the research through a synthesis of the results. The latter stands as a starting point to identify crucial areas of further research in the following section that the current researcher took into consideration to suggest the ATM. Section two describes the research questions and purposes in Phase Two. Following this section, a third section provides a conceptual framework on which the proposed ATM is theoretically grounded. In this important stage of the study, an attempt will be made to display major components that are compatible to the two crucial stages: test design and development in the operation of constructing language tests. The chapter ends with a rationale of this newly testing system. The latter will yield practical insights on the model, which is going to be implemented on the ground in Phase 3 in this study.

4.1 Synthesis of the Findings: Phase One

Phase One of the study consisted of an investigation into the diagnosis of the Current Testing System in the Algerian educational context at the micro-level. The research methods employed in this stage of the present research were four data collection methods: classroom observation, interviewing, questionnaires, and focus groups. This phase of the study answered the first research questions: what is the nature and scope of the washback effect on teachers' behaviours of EFL in the Algerian secondary schools due to the use of the Current Testing System? And, what is the nature and scope of washback effect on students' perceptions of aspects of teaching English language in the same context due to the use of the Current Testing System? These two questions also underlie other sub-questions about the

rationale and strategies utilized by both the teachers and students in relation to the prevailing teaching, learning, and testing practices in the context under exploration.

In this study, the Algerian educational context at the micro-level referred to the teachers and students in the Algerian secondary level. These subjects aimed at teaching and learning EFL for about three successive years that culminate the secondary education cycle which began earlier at the primary and middle school levels. The teachers at this level are expected to design and develop the required testing instruments that are addressed to evaluate and assess their students' attainment and progress in accordance with the taught programme and the subsequent instructional objectives set out in the relevant syllabus.

The analysis of the participants' responses generated a number of interesting findings in relation to the influence of the 'Bac' exam, as well as to the teaching and learning of English in the present context. Teachers' beliefs about the influence of the exam brought into question the validity of the claim made by the educational authorities that the 'Bac' exam has a 'positive impact' on teaching and learning. Teachers' claims were in agreement with the negative views expressed by the students according to which the 'Bac' exam and testing instruments used to assess their attainment in that foreign language has a constraining and distorting influence on the content and methodology of teaching as well as on learning.

First of all, the majority of the teachers believed that the examination influenced the content of their teaching and classroom assessment in that this examination made them pay more attention to grammar and vocabulary and some written skills rather than the oral skills throughout the school year. This practice would have encouraged teachers to use non-communicative methodology in ordinary classes, which markedly contrasted with the underlying principles of the adopted teaching methodology that prelude communicative teaching. According to the atmosphere that reigned in almost all the secondary schools in

175

Algeria, whole class activities dominated in all the class targeted to the present study. This suggests a fairly traditional teaching method where the teacher is in charge of the events in the classroom. Whole class activities mainly consist of the teacher's interaction with one or several students, very rarely students get to lead the activity. This is a common feature for all the observed classes in this study. Similarly, individual work is not very common in those classes, giving place to class time spent on whole class activities and a total dominance of the teacher talk, instead of the interaction between teacher and students, and students with students.

Teachers' reluctance to employ non-communicative methodology seemed to have on the students preferred learning strategies. The results obtained from the data collection methods regarding the students' learning strategies significantly indicated that the used strategies suggested that there was a washback effect on the learning of oral skills expected to be developed by students. This fact is evidenced in the claims made by teachers themselves that the scarcity of work on listening and speaking was due to the fact that they did not emphasize on those skills in class. In addition, evidence in the teachers' responses showed that the teaching materials used had an impact on the methods teachers used. For instance, as teachers said, one of the reasons that they did not employ communicative methodology in their classes was because of the nature of the pedagogical materials employed. Such a claim was again interpreted in terms of the influence that the 'Bac' exam exerted on those teachers in making them adopt some special teaching materials at the expense of other more effective ones. In other words, there was enough information in the teachers' responses to allow to us to see how the 'Bac' exam impacted on the materials teachers used.

Equally important was the role of the 'Bac' exam and the intense pressure it seemed to impose on teachers and students in relation to two other crucial matters that are seen to pertain to the participants' use of English inside and outside the classroom and their

perceptions of their motivation and opinions about teaching and learning. Concerning the first issue, the results showed that the participants' attitudes towards the use of that target language inside and outside the classroom was dominated by the use of some English with occasional Arabic. The high frequency of the use of a mixture of the mature and target language could be explained by the assumptions that teachers provided less practice opportunities for the students, and, in the meantime, their insistence to train these students to be fixed on some typology of activities that seem to make them to acquire the skills that could help them to respond to the questions that are recurrent in the tests they usually constructed. The above results regarding the participants' attitudes to the use of English inside and outside the classroom revealed again the pressures the 'Bac' exam exerted on the teachers and students.

Concerning the issue of the teachers' and students' motivation and opinions about instruction and learning, the exam was also said to have differential effects on these participants' attitudes. The latter commented on the influence of the exam on their psychology saying that the exam induced feelings of stress and anxiety, mainly in the students not only during its administration but also during the whole school year. Consequently, students were said to experience boredom in the classroom in addition to parents' pressure they put on them to meet the requirements of the exam that they would take later on. Those students came to look at this subject as a 'ghost' and lost any desire and motivation to enjoy the pleasure of learning this foreign language. Rather, they felt demotivated and alienated from what was going in class. Again, this was mainly attributed to the exam-oriented methodology and materials used by teachers in order to meet this exam needs.

In the same vein, another factor that might have influenced the way teachers did things in class could be attributed to certain wrong practices and misconceptions that these teachers underwent and bore. Indeed, in answer to the crucial question in this initial phase of the

177

present study which attempts to draw a transparent picture on the teachers' strategies in designing and developing achievement tests by which they usually come to test their students, it seemed that the ways the EFL teachers in Biskra region used to make their classroom ready for practice do not fit the requirements of a useful test. In precise terms, the teachers' answers showed clearly that the constructed tests do not rely on any theoretical background and do not follow any operational or procedural ways. The findings from the teachers' responses indicated that the teachers in this investigation completely ignore what to put in the crucial stages in test design and development. In other words, these teachers do not care about the purpose or purposes of the test, do not identify as clearly as possible the constructs to be measured, and totally eschew the areas related to the characteristics of students.

At another level, the same negative attitude is expressed by the teachers on the relationship between test-format and test-content in that these teachers do not adopt the divisions and sectioning of the test to the number of tasks and items that students are expected to respond. As a result, it is conspicuous to pinpoint that there exists a disequilibrium between test-tasks and test-items and the number of sections in a test. Moreover, when time comes to yield a scoring scale, what stands out clearly is that the distribution of marks per section is not consistent. One section might comprise more test-items than the other sections. Surprisingly, this loaded section with test-items might also be scored less than the other ones. In terms of the degree of difficulty of tasks, the students' answers identified that there is no gradation in designing the different questions of the test. Ostensibly, at least, the way of eliciting the items and tasks does not correspond to the common Bloom's taxonomy, which proposes that a certain hierarchy of test objectives has to be taken into consideration when ordering the different tasks. Because of this way of doing things, students find it very complicated to answer the questions.

178

Another misconception with regard to the systematic requirements of the development stage in a test significantly indicated that the teachers in this study did not know exactly what the matter of 'test specifications' is. From the analysis of findings of these teachers, it appeared that the teachers in Biskra region did not refer to this crucial step to make their tests resemble to a great extent to what they taught in class. Rather, it was clear that to develop a classroom test, the teachers picked-up ready-made tests from the available commercial books or the newspapers that propose exam papers, especially during the period of time that precedes some national examinations at the end of the school year, to give an opportunity to the students who are expected to sit for the 'Bac' exam to train themselves on how to respond to the typology of these examinations; or, they go back to the previous constructed tests by other experienced teachers and choose one of them to give to their students; or, simply they base themselves on their intuition or their experience to construct in a hurry a test and submit to their students the day of the test. Worse than this, these constructed tests are often typical examples that narrow the scope of teaching and learning since their contents usually comprise quite the same type of tasks that are often proposed in the 'Bac' exam papers. On this last point, teachers would argue that they usually proceed in this way because they are often judged on the final results their students obtain in the 'Bac' exam. On the ground of this claim, they thought it would be more practical and pragmatic for them to train their students to answer questions that resemble the ones they could find in the 'Bac' exam papers before time comes to sit for this decisive examination.

In summary, what could be retained from this synthesis of findings obtained in this Preliminary Study in this research is that the results revealed that there are many areas of influence, mainly of the 'Bac' exam, and the constructed classroom tests on the way teachers teach and learners learn in the context under exploration. Besides, they also suggested some other important factors relating to features of the local environment that are likely to have an

influence on what takes place in EFL classrooms. In precise terms, the findings confirmed that what was suggested in the first hypothesis in this research that the 'Bac' exam exerts an influence on teaching and learning of English language in the Algerian secondary schools is true. However, it is worth noting that further research on this matter is needed to explore thoroughly the claims borne by the participants in this investigation on the negative influence of the 'Bac' exam on areas discussed before, uncover some of the unanswered questions posed earlier, and identify more clearly the intensity and extent of all the factors not discussed here that seem to have also an influence on what happens in classrooms. It is hoped that this further research would manage to yield more feedback about the nature of washback in the present inquiry. Certainly, data collected from further research are expected to clarify the participants' view, as well as to bring to light new aspects of examination that are worth exploring later on.

Therefore, based on claims discussed in the first phase of the present study, it is worth mentioning that testing remains a subject that needs to be treated in the strictest care and greatest consideration. This can solely be achieved by paying importance to basic test usefulness qualities, such as validity, reliability, practicality, interactiveness, and impact. On these latter, what is recommended is to make some efforts to maximize the attainment of certain balance among those characteristics since it is argued that it is very complicated to meet the requirements of each one of these qualities individually in all occasions, instances and contexts. In this respect, this investigation in the Preliminary Study has provided us with evidence that there is a need of developing tests that correlate with their instructional objectives. There is a need of developing tests that better identify the students' specific needs and strengths, so that teachers can appropriately be informed about the students' achievement and progress. There is a need to avoid any sort of influence or authority of external examination on instruction and learning and that often leads to a negative washback. Rather,

there is a need to develop tests that must exert a positive impact on both teachers and students and make teaching more effective and learning positive.

4.2 Identification of Areas of Further Research

The discussion and synthesis of the findings obtained in Phase One in the present study indicated that there exists a number of shortcomings and contradictions in both teaching and learning of English language and test design and development of achievement tests in the Algerian secondary schools. To reiterate, this diagnosis was interpreted in terms of the strong negative washback that the 'Bac' exam exerts on the myriads of aspects in the teaching operation, and also to the exam-oriented methodology that became the current practice of EFL teachers in this context. By extension to this, it is reasonable to indicate that the negative washback effect of the 'Bac' exam mediated another influence on all the other features and factors that are intricately related to teaching and learning at the micro-level. In other words, this means that this particular exam resulted in negative consequences towards the teachers' and students' perceptions of, and attitudes towards, instruction and learning.

Given the above conclusion, it was necessary to set some guidelines that would be explored in further research in the next stages of this study, which in their turn, are expected to remedy the number of anomalies that were pinpointed earlier in Phase One of the present study. In their essence, these remedies intend to stand as practical solutions to the problems that resulted from the misunderstanding of the participants between the tasks they are hoped to fulfil and the exigencies of the terrain, the misconceptions that become ordinary with time and were accepted as a common sense in the context of investigation, and the resistance to innovation that is current among the practitioners at the micro-level and even transgresses to touch those people and domains that are integral parts of the macro-level in the Algerian educational system. Indeed, it seems that the intense pressure exerted by the educational

authorities on headmasters and teachers to obtain positive results in the final exam in the secondary school level, and the competition that the practitioners talked about to justify their focus on a certain teaching methodology have led those people to secretiveness and defensiveness of their teaching practices, approaches, and methods that are considered as the main reasons which let us obtain the results we arrived at the end of Phase One.

In light of these results and diagnosis, the researcher proceeded to suggest a road map that he developed following ideas given by writers in the field, experts, general inspectors of English language, teachers, and students who took part in this study. Added to these, the suggestions that this road map bore comprise also assumptions drawn out from the researcher's analysis to the situation under investigation due to his personal insights obtained from the results he reached, and relying on his experience as a secondary school teacher during a considerable period of time in his teaching career before graduating to the university level. For sure, all this stands as a background and a departure point that served to propose the expected remedies to the anomalies found in the Current Testing System. As a matter of fact, in some practical terms, the general guidelines of this devised road map took into consideration two fundamental objectives; first, the suggested ATM, that would replace the Current Testing System, has to be in congruence with the basic objectives of teaching and learning English language at the secondary school level. Second, the proposed model is expected to satisfy the requirements of designing and developing useful achievement tests. In this way, the intended testing model urges EFL teachers to be systematic in providing the desired tests for the specific purposes that these tests aim to achieve. More crucially, this newly, proposed testing system seeks to be an alternative to the existing system for the purpose to make testing not merely an operation of assigning marks for students, but to contribute efficiently to making teaching purposeful and learning effective.

- **What to Change?**

After reviewing the Current Testing System, a set of shortcomings and contradictions were found out. To repair these anomalies and to make the achievement tests based on the ATM useful, some modifications are recommended by the researcher in this study. These concern the following:

- We recommend a test which will be made up of three sub-tests.

- Because of the conclusion, we reached at the initial phase and taking into consideration the review provided by the participants in this study, we propose two sub-tests on testing two major language skills: listening and speaking.

- In this respect, the three sub-tests mentioned above will be as follows:

  - Sub-test One: Testing Listening.

  - Sub-test Two: Testing Speaking.

  - Sub-test Three: Testing writing and reading.

- We propose a review of the testing methods in the achievement tests based on the ATM. We propose to enlarge these tests methods to be compatible with the new sub-tests.

- Similarly, we suggest the addition of new tasks in order to make the final achievement tests more comprehensive.

- We propose the reviewing of the distribution of test-items per section and per sub-test.

- There is a need also to review the writing of instructions, make them consistent with tasks, and clear enough for students.

- We propose to avoid all the test-items that do not contribute to assess the cognitive abilities of students.

- Finally, we propose a more consistent scoring scale.

4.3 Research Questions and Purposes: Phase Two

In light of the discussion that followed in the preliminary study, it is important to note here that the present research started with two research questions that were asked and answered in Chapter 3 (section 3.1), which were RQ1 (what is the nature and scope of washback effect on teachers' behaviours of EFL in the Algerian schools due to the use of the Current testing System?), and RQ2 (what is the nature and scope of the washback effect on students' perceptions on aspects of learning English as a foreign language in the Algerian secondary schools due to the use of the Current Testing System?). In chapter 4, as the research progresses, it is likely that a new research question will be formulated and the result will be used to inform subsequent stages in the research. This research question is formulated as the following:

RQ3: "what strategies does one need to implement the ATM in the Algerian secondary schools in Biskra region?"

Phase Two of the present investigation reported in this chapter had the practical aims of gathering evidence in order to:

- design and develop an ATM for EFL classes in the secondary schools in Biskra region,

- enable EFL teachers in the context to develop achievement tests based on the newly, suggested model that can actually evaluate and assess the degree of achievement that their students are making in relation to the used syllabus and its relevant instructional objectives,

- relate language testing to language teaching,

- design achievement tests so as to encourage and enable students to perform at their highest level of performance,

- build considerations of fairness into test design,

- humanize the testing process: seek ways in which to involve students more directly in the testing process; treat those students as responsible individuals; provide them with as complete information about the entire procedure as possible, and
- draw out clear inferences from the students final scores and outcomes.

The process of formulating this basic research question (RQ3) and the relevant research purposes in this second phase in the present study were in line with what is considered by language testing specialists as competence in this area. This involves the following:

- an understanding of the fundamental consideration that must be addressed at the start of any language effort, whether this involves the development of new tests or the adaptation and selection of existing language tests;
- an understanding of the fundamental issues and concerns of the appropriate use of language tests;
- an understanding of the fundamental issues, approaches, and methods used in measurement and evaluation;
- the ability to design, develop, and evaluate and use language tests in ways that are appropriate for a given purpose, context, and students; and
- finally, the ability to read critically information about the developed tests in order to make informed decisions.

4.4 The Alternative Testing Model : A Conceptual Framework

This section presents a discussion of the framework for the ATM from planning to presenting and test administration to scoring. The basic aim of this proposed framework is to enable EFL teachers in the context of the present study to follow systematic steps that are important to design and develop a useful achievement test. In addition, understanding the

characteristics of such a framework enables those teachers to theoretically ground the process of test development, and hence avoid the current practices that most of the teachers did which clearly indicated that they did not rely on any well-defined theory or theories and did not follow systematic procedures. Whatever the situation might be, it is believed that careful planning of the test development process in testing situations is crucial because of these reasons. First, it is assumed that careful planning provides the best means for bearing that the tests will fit the purposes that they are intended to attain. Second, careful planning gives and increases more accountability for both teachers and students. Accountability implies that the teachers could say in a clear definition what was done and why. For sure, this would make the tests in terms of quality more interesting. Third, planning a test gives that feeling that the teachers or students would be satisfied because more opportunities to feel rewarded can be created.

Conceptually, the process of test design and development is organized into three stages: design, operationalization, and administration. What one has to know is that the process of developing a test is not sequential in its implementation. Indeed, in practice, it is usually stated that test development is linear, with development processing from one stage to the next. In the meantime, this process is also thought of to be iterative in the sense that the decisions that are made and the activities completed at one stage may lead teachers to reconsider and revise these decisions, and, of course, repeat activities that have been done at a previous stage. For many testing specialists, this way of conceiving things is seen to better provide a chance of monitoring the usefulness of the test throughout the development process, and therefore leads to produce the hoped, useful test. In other words, this organization avoids the common language tests that are often developed in a hurry and by intuition, and that are far from being useful.

What is significant on the testing literature is that a great deal of testing frameworks about how tests should be developed is available. Each one among these frameworks is an attempt to provide the most practical and easiest process; and, at the same time, it is a way to yield useful tests that can lead to the best final results from which inferences for decisions to be taken can be drawn out. Besides, what one has to point out on these models is that even though they present distinctions in how they are made-up, all of them share the idea that test development should begin with its initial conceptualization, referred to as the design stage, and ends up with one or more archived tests that stands as the final stage in this long process of providing tests. Thus, agreement on what components should be included in a testing framework is by no means unanimous. The frameworks proposed here are a mere adaptation to the various contexts that differ in their characteristics because of differences in the makes-up of the people targeted by the developed tests, and the purposes these tests intend to achieve. The important role of context as determinant of the way the tests are developed signifies that the latter can be meaningless if they are devoid of this context. It is then recommended that on test development matter considering the conceptual framework as essential for describing the broad parameters within which testing should fall, no matter how the suggested framework looks like or how it differs from one or the other, is something that needs to be taken with the strictest care. At the level of this point, the important question that needs to be addressed is that, "does the used conceptual framework suit the intended purposes that the developed tests can attain?"

In the present study, because there is a dire need to base the ATM on a comprehensive conceptual framework, we have thought to adopt and adapt Bachman and Palmer's (2000) model throughout this investigation. This choice is justified by a set of reasons. First, Bachman and Palmer's (ibid) model is considered by most of language testing specialists to be the recent. For many of those people, it is a synthesis to a number of available other

models. Second, this particular model is seen be to very practical and suitable to the type of tests that is explored in this study; and, it is easy to apply. Third, an important advantage this model presents is that it enables test users to review, reconsider, and re-evaluate the decisions and activities that have been taken and performed so far. This flexibility permits the adopters of this model to innovate when there is a need to do that. In short, Bachman and Palmer's (2000) model has proved to be the best choice to rely for a big number of researchers, and the present study is one of these.

After having displayed the arguments that urged the researcher to adopt Bachman and Palmer's (ibid) model as a conceptual framework for this study, below is an attempt to examine the different stages through which a language test is systematically developed. Ostensibly, the description of the different stages will take into account the main theoretical and procedural steps that one has to follow through the three stages. What is noticeable on this point is that some considerations of test usefulness have been included in these stages; this procedure seeks to emphasize that all decisions and activities involved in test development are made in order to maximize the overall usefulness of test.

4.4.1 Stage 1: Design

The product of this initial stage is often presented in the form of a document that contains some useful components in the development of test. Bachman and Palmer (ibid) insist on that in this stage a description of the components is crucial to enable test developers insure that the performance of test takers on the tasks will, to a great extent, correspond to the final outcomes. In precise terms, the product of the initial stage includes the following matters:

- A description of the purpose or purposes of the test.
- A description of the test takers for whom the test is intended.
- A description definition of the construct (s) to be assessed.

188

- A plan evaluating the qualities of usefulness; and

- An inventory of required and available resources and a plan for their allocation and management.

The purpose of this document is to provide test developers with a principled basis for developing test tasks, a blueprint, and tests. It is recommended that this document should be prepared with a great care because this can make easier to monitor the subsequent stages of development.

As indicated previously, the activities involved in the design stage should correspond to the components of the design statements. These are described briefly below:

- **Describing the purpose (s) of the test**

This activity makes clear the particular uses for which the test is intended. In these activities, the test developer has to specify clearly the inferences about language ability for language use that is intended to make on the basis upon these inferences. The result of these provide a ground for considering the potential impact of test use.

- **Describing the characteristics of test takers**

This second activity makes explicit test takers' characteristics. This concerns the identification of the population to be tested, their grade, and the topical knowledge among these tests takers. The resulting description provides another basis for considering the potential impact of the outcomes and the inferences drawn out from it.

- **Describing the construct to be assessed**

This third activity makes clear the nature of the construct to be tested. This can be done by defining the construct theoretically. By eliciting a theoretical definition of the construct to be assessed, the test developer provides the basis for considering and investigating the validity of the interpretations of the final outcomes. Also, this theoretical definition provides a ground for the development of the kinds of tasks to be used to test takers further in the

operationalization stage. Currently, the theoretical constructs in language tests are divided from a theory of language ability, syllabus specifications, or, in some cases, both.

- **Developing a plan for evaluating the qualities of tests usefulness**

This activity is about to include a plan for evaluating the qualities of test usefulness that should be a part of every stage in the test development process. The plan includes an initial consideration of the setting minimum acceptable levels for each, and a set of questions in the form of a checklist to be asked about each task developed. The feedback will be assessing usefulness in pretesting and administering and providing a wide range of information that can serve in analysing the situation. In sum, this plan for evaluating the qualities of test-usefulness includes procedures, such as a descriptive analysis of the final test scores, estimates of reliability, and appropriate analysis of the qualitative data.

- **Identifying resources and developing a plan for their allocation and management**

This last activity involved in the design stage is about the availability of the resources (human, materials, time) that will be required and that need to be available during test development. This activity aims to provide a basis for providing the potential quality of practicality of the test. Also, it seeks to monitor this quality throughout the test development process.

4.4.2 Stage 2: Operationalization

Operationalization is a practical stage. The intent of this second stage in the test development process is to yield what has been specified so far, as test constructs on the ground. In precise terms, it is about the preparation and conduct of the first actual operation of a test. Generally, this procedure can be fulfilled through the designing of what is likely named in the language-test development process as the blueprint (Weir, 1990). For Bachman and Palmer (2000), a blueprint consists of characteristics pertaining to the structure, or

overall organization of the test, along with test tasks specifications for each task type to be included in the test. In this blueprint, the test developer has to make clear the content of the test. The beginning would be the elicitation of test specifications. The latter is a form exponent to be included in the test. Ostensibly, a blueprint considerably differs from the design statement, which was displayed and elucidated earlier, in that in a blueprint there is a need to discuss and describe how actual test tasks are to be constructed, and also how these tasks are to be arranged to form the test. In such a procedure, each task has to be well differentiated from the other existing tasks. In addition to this, another part that needs to be evaluated in this stage is to elicit a set of test methods from which a test developer is expected to select some to adopt in the developed test. Customarily, this is a very crucial procedure since it stands as a background that enables to vary the test methods and helps in avoiding to be fixed on the same testing activities and hence allows a better judgment of the test takers' performances; what is required here is to devise appropriate test methods that are usually relevant to the intended purposes a test developer is supposed to attain.

Next, once the elaboration of the content of the test is finished, the test developer should turn to the issue of test framework. This specific matter includes defining as clearly as possible the structure of the test, make explicit the number of parts in one test, identify the number of tasks per test, and finally, indicate in precise terms the time allocated for each task. Doing these, the following step would be to write test instructions. This procedure involves describing fully and explicitly the relationship between each task and its correspondent instruction; what is usually specific on this point is that some instructions are general and apply to the test as a whole, whereas others are closely very linked with specific test tasks. The primary purpose of test instruction is to insure that test takers understand fully the exact nature of the testing procedure, and how they are going to respond to these tasks. The essential components of test instructions, therefore, are:

- Statement of the purpose (s) for which the test is intended,

- Statement of the abilities or constructs that the test is intended to measure, and

- Specification of the procedures and tasks.

If instructions are to accomplish their purposes, then there is a need to make them understood by the test takers. In deciding how best to make the instructions understandable, some criteria must be paid some attention. These are:

- The language,

- The channel through which the instructions are presented,

- The need for providing examples, and

- Also the need to pre-test the instructions with test takers.

About the kind of language these instructions one should bear, it is essential to take into consideration the amount of difficulty of this language. What is recommended at this level is that the more the language is clear and straightforward, the less ambiguous and confusing the instructions will be. In this sense, test takers are expected to understand what they are asked to do from just reading the instructions and not spending their time in decoding the questions posed. Here, understandability can be facilitated by presenting the instructions in a channel that is most appropriate to the purpose and abilities to be answered, and that test takers are most likely to find easy. This means that in tests whose input is presented in the normal channel, instructions are also typically presented in this way. To make this last assumption clear enough, let us consider instructions for completing a test with tasks that require typically to be presented in writing, it is crucial to adopt the channel that is based on composing and only that. However, in tests with listening or speaking activities, which typically involve input in aural channel, instructions would be presented in this channel.

The final step in this phase is to devise a scoring method. This process consists in establishing a method for quantifying responses for test tasks. It involves two fundamental

operations. First, it is important to identify the criteria by which the quality of test takers' responses will be identified. Second, there is a need to determine the procedure that will be followed to arrive at a score. Certainly, these two steps are thus essential to reach a measure to any qualitative, descriptive information from the responses of the test takers. Because the measure that test developers obtain are seen to be essential in assisting to making decisions about test takers, the method used to arrive at the scores is also crucial in the whole testing process.

In determining what a scoring method to use, two aspects of test development, discussed previously need to be considered. These are the theoretical definition of the construct to be measured, and the other one is the test-task specifications. For the former, the way one defines the construct for a particular situation will determine which areas of language ability that has to be scored, and the role of typical knowledge in the scoring method. Besides, the construct definition will also determine the type of score to be reported, whether this is a profile for scores for different areas of language ability, a single composite score, or both. For the latter, it is found that how the way the test-task is specified to define the type of the intended response has a clear implication to the final score. Thus, tasks intended to elicit a selected response can generally be scored objectively, whereas if the intended response is limited, it is necessary to consider the nature of the language of the intended response in determining the specific scoring method to be used.

4.4.3 Stage 3: Administration

The administration stage of test development involves submitting the test-to-test takers, collecting information, and analysing this information for two purposes:

1. Assessing the usefulness of the test, and
2. Making the inferences or decision for which the test is intended.

To accomplish these two purposes at the final stage in the test development process, a test developer should describe the procedures that s/he is expected to follow for administering a test, and, then, has also to explain aspects that are related to this stage. Concerning the procedures for administering a test, the first step is to prepare the testing environment to be consistent with the suggested test specifications in the test blueprint. This involves arranging the place of testing, or setting, materials and equipment, personnel, time of testing, and physical conditions under which the test is administered. To provide an example of this aspect, for instance, if the test requires testing the speaking skill, the design specifications need to choose a large place to allow the test takers to perform the test in a suitable environment. The preparation of the materials appropriate to testing this skill requires an organized choice of the technical instruments. In short, behind this first step is to maximize the usefulness of the test by making the physical conditions under which the test is administered appropriate for the purpose it intends.

The next step is to maintain a supporting testing environment throughout the test. This includes avoiding distraction due to some external factors, such as temperature, noise, excessive movement and others. It is important for the test developer to think through what kind of assistance s/he wants to be provided during the test before administering it. Much of the time, this assistance focuses more on the test-taking environment to control variability in order to enhance reliability. This means that equal consideration to all test takers should be given in order to conduct them to perform at their best; in the meantime, this makes them avoid discontent before being tested. A fact that often leads to negative influences on the test takers' final outcomes. In this respect, setting a supportive test taking environment needs to be carried out with all qualities of test usefulness. Again, this could be realized only by thinking through what kind of assistance one needs to provide to control the set of variables in the setting in order to enhance the degree of reliability. The final step in administration is

collecting the tests. For many testing specialists, the behaviour preceding the act of collecting the tests is important since if it is done in a way that does not harm the test takers' feelings, it can be a positive contribution to the overall usefulness of the test. For instance, as a good behaviour, instead of hurrying to collect the tests, the test developer might give test takers a chance to write down a few comments on their experience. While test takers are providing the feedback, the test developer can quietly collect the answer sheets. This simple behaviour can greatly contribute in making the test takers respond the questions of the tests at their best. This directly leads to contribute to the whole usefulness of the test.

About the aspects of test administration, test developers usually distinguish between pre-testing and operational testing; pre-testing is to collect information about test usefulness in order to make revisions in the test itself and in the procedures for administering it, rather to make inferences about test takers. The purpose on carrying out this step is to gain mostly qualitative feedback. This feedback can show areas of problems in relation to tasks, instructions, and administrative procedures. Following this, the next step in the pre-testing cycle is usually a field trial in which the test is administered under operational conditions. The purpose of a field trial is to collect data for assessing usefulness, and not to take inferences about test takers. What is noticeable about these activities is that, in some cases, they are reduced or completely omitted. It is the case with classroom tests where in such procedures could be limited because of the nature of this type of tests. However, it is advisable that submitting the test to students or follow teachers in advance is a good idea since this can provide useful information to improve it.

Because one of the major aims of the present research is to sensitize EFL teachers in the context under exploration that the issue of test development is a very systematic process, it is essential that those practitioners have to go through an entire operation of creating and using the tests they often develop. To reiterate, this particular process is organized into three

195

stages that were in some detail explained in this section. What should be kept from this description is that the latter serves as a theoretical background on which the suggested ATM in this study rests. In other terms, the conceptual framework that has been elaborated so far would guide teachers of English language in the Algerian secondary schools develop useful achievement tests that themselves could reach valid and reliable final outcomes. Following this theoretical presentation of the ATM, the next section is going to display this model in its practical manifestations. Of course, the intention of this sectioning is expected to urge EFL teachers to be systematic, and give them an opportunity to make, consider, and revise the decisions they have taken in one stage or another.

4.5 A Rationale for the Alternative Testing Model

This section describes the suggested ATM for EFL students in the Algerian secondary schools in Biskra region. As has been discussed is section 4.4, the present section is divided into two sub-sections. Sub-section One presents the major elements that are an integral part to the design stage. In the second sub-section, practical considerations on the ATM will be identified.

4.5.1 Test Design

4.5.1.1 The Purposes of the test

The ATM aims at helping teachers to:

- Understand how students are progressing in learning the content of the programme devised for them in achieving the instructional objectives. In other words, teachers have to see the degree to which their students meet minimum standards of mastery of the content of syllabus, and the relevant objectives;

- Find out more about students' weaknesses and strengths;

- use information from the assessment as a new basis for future lesson planning and preparation;

- provide evidence of the results of instruction and learning, and hence feedback on the effectiveness of the teaching programme. Therefore, teachers have to proceed to making appropriate decisions. Decisions about instruction include determining what parts of unit/theme have been taught and which parts require review;

- provide information relevant to making decision about students such as what specific kinds of learning materials and activities should be given to them.

- provide systematic feedback for the teacher. s/he can see how well the students did on the material tested and check any discrepancies between areas of instruction or review;

- clarify instructional objectives and their relevance to instructional materials and activities of students following the programme of instruction; and

- assess the effect of the programme of instruction with regard to students' development in the ability to communicate appropriately and efficiently and their ability in exploring and exploiting materials related to instruction.

4.5.1.2 Characteristics of test takers

1. Personal characteristics

EFL students, subject of this investigation, are both males and females. Their age is between 15 and 17. They come from different socio-economic background. Their native language is Arabic. English for them is the second foreign language after French. They have started studying English language since the first year in the middle school. At the end of their secondary school studying cycle, they are expected to have spent seven years studying that language. They are also supposed to be familiar with achievement tests. What is noticeable

197

about this matter is that the contents of these tests might vary from one class to another, depending on the contents of the programme taught by teachers.

2. Topical knowledge

It is fair to mention that those students have not a special topical knowledge. What is significant on this point is that the topical knowledge includes the same course content with all its prevailing language, structures, and vocabulary. The students in this context have pursued the same academic studying cycle; they have been taught with the same syllabus, approaches methods, techniques, and they used the same textbooks.

3. General level and profile of language ability

Having got the same socio-cultural background, living in the same environment, and studying using the same instruments do not obligatory mean that these students have the same language proficiency. The levels and abilities of those students can usually be tailored by other mental and affective characteristics. Generally, these concern the cognitive capacity and psychological make-up of each student. Certainly, these two features always affect how a student grasps and processes the amount of knowledge s/he is presented with, and, of course, it leads to different learning styles. In precise terms, you may sometimes find within the same group of students in a given class different levels of achievements. On one hand, students with a high aptitude to learning a language are usually able to have a good mastery of the subject learnt in a quite short time and with an ascendant progress. On the other hand, another category of students with a low aptitude to learning that often hampers them to attain a better achievement level.

4.5.1.3 Definition of Constructs to be Measured

For this ATM, a syllabus-based construct definition is used. The constructs to be measured are specified in terms of the target learning objectives that are set out in the syllabus. In broad senses, the general aim of the learning objectives set out in the syllabus for

EFL learners in the Algerian secondary schools is to reinforce knowledge and skills following the four-year course received at the middle school. In addition, these instructional objectives seek to develop basic competencies through various activities that EFL learners at this stage should develop. These competencies are determined in precise terms in the syllabus as follows:

- Interact orally in English

- Interpret oral and written messages

- Produce oral and written messages

To make these instructional objectives more explicit, they are categorized as to enable EFL learners achieve communication in its various forms, aspects, and dimensions. Besides, to developing these cited competencies, the objectives could be achieved through the development of the learners' mental abilities because they are seen to be fundamental for any acquisition of knowledge. Those latter are stated as follows:

- Knowledge to be able to state - recall - reproduce.

- Comprehension to be able to predict - identify - explain - illustrate.

- Application to be able to predict - select - use - construct.

- Synthesis to be able to summarize - argue - organize - conclude.

- Evaluation to be able to judge - select - support - attack - evaluate.

In this sense, students are expected to master the basic characteristics of language, which are:

- Listening to and understanding oral message.

- Expressing oneself with a certain fluency so as to be understood.

- Reading and understanding of various kinds of authentic texts.

- Mastering different writing skills and writing about major features of this language.

199

4.5.2 Test Development

4.5.2.1 Test Contents

4.5.2.1.1 Test Specifications

For the ATM, a syllabus-based construct definition is used; this includes specifying language contents and exponent that were presented to students in lessons. Examples of these are:

1. Functions

2. Skills

3. Grammar

4. Vocabulary

For convenience, the researcher proceeds to display an example of test specifications to develop achievement test for second year EFL classes in the Algerian secondary school. The choice of this level among the secondary cycle is interpreted by the fact that it corresponds to the population targeted in this research. However, what is worth mentioning on this point is that what is applicable on this level can also be applicable to the other two remaining levels, which are the first and the third year levels.

A review of the contents of the second year syllabus shows that the latter consists of units. They are organized topically as in all secondary school textbooks (see Table 4.1). The students are expected to prepare projects after every unit. The project is the final task. It requires the application of both types of cognitive skills: 'lower order' skills, which imply acquiring new knowledge, understanding new facts, and ideas and applying them to solve problems; and, on the other side, 'higher order' skills that are about analysing information by breaking it into small parts and patterns and evaluating new information by forming opinions and judging the quality of that new information.

Unit	Branches			
	Technology (Maths) Experimental Sciences	Accountancy, Economy	Letters, Philosophy	Foreign Languages
Unit 1: Diversity			+	+
Unit 2: Peace and Conflict Resolution	+	+	+	+
Unit 3: Poverty and World Resources	+	+	+	+
Unit 4: Technology and Innovation	+			+
Unit 5: Fiction or Reality	-		+	+
Unit 6: Disaster and Safety	+	+	+	+
Unit 7: Management and Safety		+		

Table 4.1: Units and their distribution per branch.

Each unit consists of five main parts:

1. Part One: describing language (constituents of language).

2. Part Two: developing skills/build basic language skills and intellectual skills.

3. Part Three: putting things together (project itself).

4. Part Four: where do we go from here? (Self-assessment).

5. Part Five: exploring matters further exclusive reading, acquisition of

vocabulary and consolidation of grammar).

4.5.2.2 Test Methods

The main proviso for testing within the ATM is that test methods as far as possible reflect the range of contributory skills and content coverage that are included in the designed test specifications. It is also crucial the developed achievement tests should have a strong

washback on practice in the classroom. In this sphere, test methods are used to construct tests but are not in themselves tests. There is some evidence in the literature (see Weir, 1988; Bachman & Palmer, 2000) that test methods if chosen in the right way when developing a language test could make such a test valid and reliable, and hence affect the test takers' final outcomes positively.

In what follows, a list of very useful test methods is proposed that EFL teachers who adopt the ATM can refer to whenever they come to develop achievement tests for their classes. It should be made clear that the process of addition of test methods has to remain open and flexible when the test developer thinks that there is a need for that. All what one has to consider is that the choice of any test method should be congruent to what will be tested and serves the appropriate assessment that the test itself intends to attain. Thus, the test methods to be integrated in the ATM are as follows:

1. Multiple choice	7. Gap filling
2. Dichotomous items	8. Selecting deletion gap
3. Matching	9. Cloze
4. Information transfer	10. Dictation
5. Listening recall	11. Editing task
6. Ordering tasks	12. Controlled writing tasks

4.5.2.3 Testing the Major Macro-skills and Language Components

4.5.2.3.1 Testing Listening

Testing listening is testing the learners' ability to comprehend any kind of discourse, and, in the meantime to discriminate phonemes and the mention of what has been heard. However, on these abilities two different views are identified. First, a view that focuses more on discrete listening items, such as sound recognition, intonation, and stress through the use of short items. The sum of these discrete sub-skills should be equivalent to the learners' proficiency in listening. Second, another view seeks to differentiate between tests of listening

auditory discrimination and decontextualization test of listening. The shift from the first view to the second is characterized by more emphasis on meaning and communication rather than a mere concentration on discrete and isolated items of discourse. Therefore, in regard to these two isolated different approaches to testing listening, in the ATM, it is believed that associating the two views could be of a great usefulness for both teachers and students to obtain insightful feedback for the process of assessing this skill. Indeed, approaching the assessment of listening in this way provides more inferential final outcomes. At the level of test development, this enables to maximize the major two test-qualities validity, and reliability. In this respect, in sub-test one in the ATM, two sections about testing listening are proposed, section one deals with testing listening in context, whereas section two is about testing only discrete listening items. This procedure aims at yielding useful achievement tests that repair the absence of testing listening in the Current Testing System.

1. Objectives

In the ATM, the fundamental objectives of testing listening are to enable test takers to recognize without too much effort the following:

- sound patterns, intonation patterns, stress, and word grouping;

- grammatical sequences;

- modifiers, and

- hesitation expressions;

- understand different types of aural messages linked to everyday activities, such as face-to-face conversation, telephone conversations, and recorded conversations, and;

- finally, understand the main basics of a talk delivered in Standard English on various cultural and/or social themes. This talk would be worked out in terms of texts, grammar, overall organization, and other possible interpretations.

2. Type of texts

To test listening, the ATM proposes to choose texts from authentic materials consisting of utterances with a high probability of use by native speakers of English language. All what is recommended is that these materials should correspond to the overall language proficiency of test takers. For instance, utterances and speech should be delivered at normal speed with standard language, and preferably that these utterances and speech contain a number of pauses. With the progress of lessons and the school year, the level of these texts could advance to judge how much students have improved. Anyway, what is essential is that the materials included in testing listening should respect the standard norms of the language taught in class.

3. Possible methods

In line with the suggestions elicited in the ATM, the possible methods proposed to testing listening are equally applicable for using their formats to assessing that skill. Examples of these methods are described below:

   i. Multiple Choice Questions (MCQs):

In testing listening, this test method could be effective in that:

- The marking process is totally objective.

- Agreement is usually reached as to the correct answer for each item.

- The marking is simple and rapid and often more cost effective.

- It is less time consuming and demanding.

- It is very useful with testing extensive listening skills. At this level, the main objective of a listening test is to evaluate the test takers' comprehension.

ii. Short Answer Questions (SAQs):

Contrary to the previous test method, a short answer test method is a technique that requires test takers to write down specific answers in spaces provided on the test paper. The test method is extremely useful for testing listening because:

- Answers are not provided for the test takers as in multiple choice; hence, if a test taker gets that answer right, one is more certain that this has not occurred for reasons other than comprehension of the test.

- With a careful formulation of the questions, a test taker's response can be brief and thus a large number of questions may be set in this format, enabling a wide coverage.

- If the number of acceptable answers to a question is limited, it is possible to give fairly precise instructions to the test developer who marks them.

iii. Information Transfer

This test method requires from the test taker to transform a verbal information into a non-verbal form, for example, by labelling a diagram, completing a chart, or numbering a sequence of events. The merits of this test method are as the following:

- A particular advantage for using information transfer in assessing listening is that test takers do not have to process written questions while trying to make sense of the spoken input. It is particularly efficient for testing an understanding of sequence, process, relationships in a test and classification.

iv. Dictation

This test method involves test takers listening to dictated material that incorporates oral passages typical of those they might encounter in target language situations. The advantages of dictation in assessing the listening skill are summarized in these points:

- A dictation can provide more reliability to the listening test through a large number of items that can be generated, as well as being valid for specific strategies where dictation might feature as a target group activity.

- It is considered to be effective in listening since it reflects more faithfully how test takers process language in real-life contexts. Also, it is a welcomed test method in assessing this skill because as a testing device it measures two different language features to be useful in providing means of assessing any one particular skill.

- An important factor in incorporating this test method in assessing listening is because it is an effective device in testing listening comprehension since it assesses a broad range of integrative skills and sub-skills.

- This test method enables test takers not only to discriminate phonological units, but also their ability to make decisions about word boundaries: in this way, a test taker discovers sequences of words and phrases that make sense and from these s/he reconstructs a message.

v. A listening recall

The premise of this test method is that the test taker is given a passage from which certain words have been omitted. The words deleted are normally content words felt to be important to an understanding of the discourse and the blanks occur at increasingly frequent intervals. Test takers are given a short time to listen and read over the text; they have to fill in the blanks. The advantages of this test method are:

- Like dictation, it can be administered rapidly and scored objectively, and also it allows the test taker to focus on test items which are deemed to be important.

- It is very applicable to testing a large number of students since it is easy to construct, administer, and score.

Besides, other test methods that can be incorporated in assessing listening include examples, such as:

- Note taking

- Paraphrasing

- Choosing a title

- and others.

4.5.2.3.2 Testing Speaking

The introduction of a speaking test seems the most notable feature of the ATM. What is significant on this particular skill is that all of the teachers in the context under study but one used English as their means of classroom communication in English lessons, but they rarely focused on to develop the speaking abilities of their students. For most of these teachers, it was generally not felt important to do specific work on speaking since it was not part of the Current Testing System and none of the students needed to take the test of this skill. This

belief has a great impact on their students' performance in this ability in particular and English in general. The challenge set out in this proposed new testing system in this study is to incorporate a sub-test testing speaking in order to repair the existing anomalies.

A basic assumption made in the ATM is that the assessment of the speaking skill aims the development of the ability to interact successfully and efficiently in the target language, and that this involves both comprehension as well as production. These two features should offer plenty of hope for meeting the objectives for testing this important skill. Following the pattern of the previous sub-test, we will deal with the same points in turn below. The beginning would be an elucidation of the objectives that the testing of the speaking skill in the ATM target:

1. Objectives

In the ATM, the categorization of the objectives of assessing speaking relies on three fundamental types. These types of purposes are:

a. Informational purposes: Test takers should be able to:

- provide personal information;
- provide non-personal information;
- describe sequence of events;
- make comparisons;
- give explanations;
- present an argument;
- provide required information;
- express need;
- elicit help;
- seek permission;
- apologize;

- elaborate an idea;
- express opinions;
- justify opinions;
- complain;
- analyse;
- make excuse;
- paraphrase;
- summarize what they have said;
- express references;
- make comments; and
- indicate attitude.

b. Interactional purposes: Test takers should be able to:

- express purpose;
- recognize other speakers' purpose;
- express agreement;
- express disagreement;
- elicit opinions;
- question assertions made by other speakers;
- modify statements or opinions of other speakers;

- attempt to persuade others;
- check that they understand or have been understood correctly;
- elicit clarification;
- correct themselves or others;
- indicate understanding (or failure to understand); and
- indicate uncertainty.

c. Purposes in managing interaction: Test takers should be able to:

- initiate interactions;
- change the topic of interaction;
- share the responsibility for the development of an interaction;

- take their turn in an interaction;
- give turns to other speakers;
- come to a decision; and
- end an interaction

2. Types of Materials

As said previously in the development stage in the test specifications form, it is important to sample the contents of what has been taught in class and is suggested in the relevant syllabus. Thus, the type of materials to testing this skill has to do more with tasks, such as:

- Presentation (monologue)
- Discussion
- Conversation
- Interview

In line with these tasks, the selected topics should correspond to some particular aspects, such as:

- Background/Nature: the topics should be familiar and interesting to the test takers.

- Language: the kind of language has to be the same as the one known as the standard British English (this is what the syllabus recommends)

- Accent: it should be Received Pronunciation, or as it is also known as RP.

- Style: commonly the style should be formal. However, in some cases, it is also advised to choose some topics from informal language. This makes students acquainted to different language styles.

- Vocabulary range: generally the diction should be taken from what the students studied in their class. Some vocabulary-items on the technical jargon are preferably to be avoided.

- Rate of speech: this rate has to vary according to the task and topic presented in the test.

What is more, specifying the objectives into three different categories (Informational, interactional, and management), as it is the case above, should help in creating tasks which will elicit a representative sample of each. It is indicated that the greater the detail in specifications of content is, the more valid the test is likely to be. In this context, teachers may wish to select any set of specifications for their own purposes.

3. Possible methods:

Having established what it is that needs to be tested, there is available a range of test methods of varying types. In what is next, it will be a review of some potentially valid test methods and techniques to test speaking that are integrated in the ATM.

a. Verbal essay

Following this test method, the test taker in expected to speak for a short time (often this time is pre-arranged by the test developer according to the time allocated for the whole test), on either one or more specified topics. The merit of this test method is:

- The test taker has to speak at length which enables a wide range of criteria including fluency to be applied to the output. More discrete short questions or pointed situations to which the test taker has to respond often severely limit the range of criteria that are applied.

b. Oral presentation

The test taker is expected to give a short talk on a topic s/he has either been asked to prepare beforehand or has been informed shortly before the test. The difference between this test method and the verbal essay is that in the former the test taker is allowed to prepare for the task, whereas in the latter s/he is not. Its advantage is that:

- It is often very effective to get the test taker to talk about a topic s/he knows in advance what it is about. Experience has shown that in the case of this test method even though time is very short, the test taker finds it easier to talk and express his/her ideas. What is also important in assessing speaking here is that this technique permits to elicit a sufficient sample of the test takers' speech for sensitive assessments to be made.

- By integrating the task with the previously proposed tasks in class, the oral task can be made to equate realistically with real-life tasks that the test taker might have to perform in the target situation.

c. Information transfer: description of a picture sequence

The test taker sees a panel of pictures depicting a chronologically ordered sequence of events and has to tell the story. The test taker is given some time to study the pictures. Some of the advantages of this test method are:

- The task required of the test taker is clear. It does not require him/her to read or listen and thereby avoids the criticism of contamination of measurement provided by the pictures that are not culturally or educationally biased.

- It can be very efficient to enable the test taker to provide an extended sample of corrected speech which allows the application of a wide range of criteria in assessment. It is also useful in making the test taker use some language components like the tense.

- This test method is valuable because it relies on the pictures since it enables test takers to show their mastery of the target language.

d. Interaction tasks: information gap

In this particular test method, the test takers work in pairs and each is given only part of the information necessary for completion of the task. They have to complete the task by getting missing information from each other. Test takers must communicate to fill in an information gap in a meaningful situation. The merit of this test method is that:

- It is a good example where the test taker is expressed to communicate and hence show how much s/he masters the language. It aims to assess appropriately the performance of this test taker.

- As a normal feature of the interaction, a test taker can use question forms, elicit information, make requests, ask for clarification, and paraphrase in order to succeed in the task.

4.5.2.3.3 Testing Reading

Because it is a receptive skill, the testing of reading bears quite the same characteristics as testing listening. It is due to this similarity that what is generally proposed as test methods in testing listening can be used to test reading. Crucially, through the reading skill, students are expected to be able to:

- recognize the structural clues such as clauses, words, transitions, and tenses.

- distinguish word grouping and relationships with other word groupings (sentences, part of sentences, ...), and

- anticipate what probably follows while holding in our intermediate memory elements from what precedes.

The information students often derive from reading are various. Currently, these are as the following:

- Grapho-phonic information i.e., sound-symbol relationship

- Syntactic information derived from structures.

- Semantic information i.e., content words

- Socio-cultural information or information derived from specific cultural connotations.

1. Objectives

In the ATM, the objectives of testing reading are directly related to the students' needs or the course that are identified in the syllabus. Commonly, these set out reading objectives may be thought of as:

- scanning the text to locate specific information,

- skimming the text to obtain the gist,

- reading and showing understanding of views of various kinds of authentic text (narrative, descriptive, argumentative) of intermediate difficulty,

- identifying major ideas, sub-points, and supporting details,

- recognizing the rhetorical devices that typically go with difficult types of development ideas in writing, and

- recognizing structural patterns typical of specific types of writing.

2. Types

In the ATM, the chosen supporting texts to be included in the part of the test testing reading should obey some specific features. These are as the following:

a. Types of texts

The type of texts must take into account the typology of texts presented to students in instruction. Students need to be familiar with these texts, but they should not be a copy of them.

b. Sources of the texts

An important criterion when time comes to assess the reading skill is to choose the supporting texts. What is suggested in the ATM is that the sources of texts must be varied. Test developers can refer in addition to the official textbooks, to newspapers, magazines, academic journals, letters, internet, and the like. Crucially, it is advisable to opt for supporting texts that are intended for a native-speaking audience. Meanwhile, these texts should not go beyond the general proficiency levels of test takers because the type of texts that is developed in this new testing system is meant to evaluate students' achievement in English language, and not their proficiency. An additional remark that is hopefully to be

considered by a test developer concerns to mention the source or sources of the text if it is integrally taken or adapted from an external source.

c. Degree of difficulty of texts

In the ATM, a common feature about the supporting texts is that they should be of an intermediate difficulty. It goes without saying that in the case of this newly, introduced testing system, the adopted texts in assessing reading need to correlate with the common overall level of test takers. More importantly, test developers should not choose texts of particular kind simply because they are readily available.

d. Length of texts

In the ATM, the length of the supporting texts must take into account the time allocated for the test, the number of test-items, and the general level of the test takers that will be assessed. Generally, the model proposes supporting texts of a number of words that varies between 170 and 200.

e. Language complexity

In the ATM, assuming that it is only the reading ability which is going to be tested, test developers should not chose texts with a high degree of complexity. Instead, it is more appropriate to select supporting texts that are accessible to the test takers. This does not obligatorily imply that these test developers have to choose texts that students have already read or in close approximation to them. Besides, in this model, it is preferably to opt for texts that have not be culturally laden. The texts should not excite or disturb test takers. Rather, they have to be part of their interests and relevant to what they have learnt in class.

3. Possible methods

The test methods, which are thought to be appropriate testing reading, include examples, such as:

    a. Multiple-choice Questions (MCQs)

As has been mentioned so far in the part devoted to testing listening, this test method is also applicable to assessing reading. Some additional merits of this technique could be:

- With this test method, an almost complete marker reliability is reachable. The test takers' scores, unlike those in the subjective formats, cannot be affected by the personal judgement or idiosyncrasies of the teacher. The marking, as well as being reliable, is simple, more rapid, and more cost effective than other focus of written tests.

- Because items can be pre-tested fairly easily. It is usually possible to estimate in advance the difficulty level of each item and that of the test as a whole.

- The format of the multiple-choice questions test method makes it easy for test takers what is required of them. In open-ended format ambiguities in the wording of questions may sometimes lead to the test takers submitting answers to questions different from those which the teacher intended to ask.

- In some other test methods such as short questions or questions that have open-ended formats, the test taker has to deploy the skill of writing. The extent to which this affects accurate measurement of the trait being assessed has not been established.

The multiple-choice questions test method avoids this particular difficulty.

b. Short answer questions

This test method bears the same principles that were described in the part talking about the same technique testing the listening skill. In addition to what has been said under this skill, short answer questions are very useful to assess the reading ability. In reference to this, most advantages of the use of this technique in this specific domain can be summarized as:

- Activities such as inference, recognition of a sequence, comparison and establishing the main idea of a text, require the relating of sentences in a text with other items which may be distant in the text. This can be done effectively through short answer questions where the answer has to be sought rather than being one of those provided.

- More importantly, this test method in this area can also provide more reliable data about test takers' reading ability.

c. Cloze

It is described as the test technique where in words are deleted from a text after allowing a few sentences of introduction. The deletion rate is mechanically set, usually between every fifth and eleventh word. Test takers have to fill each gap by supplying the word they think has been deleted. About this test method, there is a good deal of supportive evidence in the literature for using the cloze format. Some of its merits are:

- This test method is easy to construct and easily marked if the exact word scoring procedure is adopted.

217

- This technique of testing reading is considered to present a high degree of consistency since with a fifth word deletion rate a large number of items can be set on a relatively short time.

- In the literature, cloze tests are often regarded as valid and uniform measures of reading comprehension.

d. Selection deletion gap filling

To overcome some negative findings due to the mechanical deletion cloze, selection deletion gap filling has been proposed. In its essence, this test method consists in selecting items for deletion based upon what is known about language, about difficulty in text, and about the way language works in a particular text. Linguistic reasoning is used to decide on deletions. The advantages of this method are as follows:

- Selective deletion enables the test developer to determine where deletions are to be made and focus on those items which have been selected a priori as being important to a particular target audience.

- It is also easy for the test developer to make any alternation shown to be necessary after item analysis and maintain the required number of items; this might involve eliminating the items that have not performed satisfactorily in terms of discrimination and facility value.

e. Information transfer

In testing listening, we have referred to this test method as one of the welcomed techniques to be used. Similarly, in testing reading, this type of test methods stands as one of the most useful. Information transfer consists in transforming information from a test through

reading to label a diagramme, complete a chart, number a sequence of events, or other tasks; two main merits of this test method are elucidated below:

- This test method is suitable for understanding a process or classification on narrative sequence. Also, it is effective in testing a variety of other text types.

- It is realistic and its authenticity gives it a high degree of validity in these contexts.

4.5.2.3.4 Testing Writing

Writing is the act of the expression of ideas in a consecutive way according to the various conventions of the language; the major concern of developing writing is to achieve accuracy. To realize this, students need to handle a set of tasks. First, they have to get an understanding of the whole system of the target language. Second, they should be successful in choosing the suitable mechanics devices, mainly punctuation, and know how to use it appropriately. Next, it is crucial that students get a good control of the language exponents, so that they manage to produce a piece of writing comprehensive for those who read it. Finally, they should be able to select from a range of possible combinations of words those ones that bear the right sense. Hence, they should be sure that these selected diction and lexical items belong to the appropriate register.

1. **Objectives**

In the ATM, two fundamental objectives of testing writing are described as:

- To measure the adequacy of expressing some information received from what test takers have learnt in their lessons through the exploration of the various didactic units and their relevant sub-themes,

- To judge the extent to which test takers are able to provide an accurate piece of writing taking into account aspects, such as: the information required, the grammatical accuracy, the appropriate lexical choice, the way how ideas are arranged, and the general idiomatricity i.e., the feeling for authenticity.

2. Possible methods

In the Alternative testing Model, two approaches are proposed. Firstly, writing can be divided into discrete levels e.g., grammar, vocabulary, spelling, and punctuation, which themselves are to be assessed separately through objective tests. Secondly, more direct extended writing tasks of various types could be constructed. These would have greater construct, contents and face validity, but would require a more objective assessment.

Below is going to be a short account of some test methods incorporated in the present testing system that are a combination of the different approaches. Some of them are:

    a. Editing tasks

In this test method, the test taker is given a text containing a set of errors of grammar, spelling, and punctuation of the type noted as common by remedial teachers of students in the target group, and is asked to rewrite the passage making all necessary corrections. Some of the advantages of these test methods are:

- As well as being a more objective measurement of competence, this task may have a good washback effect in that test takers may be encouraged to elicit their written productions more carefully.
- It is more face valid than other indirect techniques as it equals part of the writing process.

220

b. Composition

This test method is very traditional for getting test takers produce a piece of writing. It varies from a limited number of words to several sentences. The particularity of this test method is that the test taker is not guided or controlled in any way as to how s/he is expected to answer the question. Some of the merits of this test method are:

- The topics are extremely easy to set. Also, it is a familiar test method to both the test developer and test taker.

- It is suitable for testing sub-skills such as the ability to develop an extended argument in a logical manner, which cannot be tested in other ways.

- The big advantage it shares with other test methods of extended writing is that the sample of writing is produced which can provide a tangible point of reference for comparison in the future.

c. Controlled composition

In this test method, test takers are engineered to perform a given task by composing some ideas that are often suggested to guide test takers of developing the topic. These are various types of stimuli that can be used in controlled writing tasks. Stimuli can be written or even non-verbal, such as a graph, a plan, or a drawing, which the test taker is asked to interpret in writing. Some of the advantages of controlled composition are:

- The information is clear and precise that the test taker does not have to spend a long period of time decoding a written test.

- The task can be effective when the test taker is asked to comment on particular topics shown in a graph, or to compare and contrast one set of figures with another.

- Different manners can be used to elicit a composition of different language functions; such as argumentation, description of a process, comparison and contrast, or writing a set of instruction.

4.5.2.3.5 Testing Grammar and vocabulary

A. Testing Grammar

In the ATM, testing grammar is separate from the other language components. The main reason of this strategy is to ensure a high validity of the test since testing this language exponent in this way would give more chance to assess a big amount of grammatical structures dealt with in instruction.

1. **Objectives**

The major aim of testing grammar is to enable teachers to measure the mastery of this language component taught in class. Also, it is a source of useful feedback as to the adequacy of the teaching methods used by these teachers. Such feedback can stand as a ground on which to make inferences on how to employ appropriate remedial work to make the language learnt by students very accurate.

2. **Possible methods**

The test methods to be deployed in this testing system on testing grammar aim to allow test takers to supply grammatical structures and recognize their correct use. Some of these methods are displayed below:

- Paraphrasing
- Combining statements
- Making the questions
- Expressing it differently
- Supplying the appropriate tense
- Spotting the mistakes and correcting them

B. Testing Vocabulary

In the Alternative testing model, testing vocabulary is paid a great attention. This attention rests on an approach that is meant to make it possible for the teachers to demonstrate accurately the development of the amount and quality of vocabulary stock students have acquired during a period of time. In this approach, there is a need to specify the kind of vocabulary to be tested, and suggest adequate test methods that ought to make the achievement of the objectives set out by the teachers accessible.

1. Objectives

Since the main concern of the ATM is to measure the extent of achievement of English language in accordance with the contents of the taught syllabus, and the relevant instructional objectives, the objectives of testing vocabulary in this testing system are categorized as follows:

- It would be a fair proof that vocabulary learning is taking through a systematic approach and an efficient method.
- It would be possible that students' comprehension of that foreign language is deepening through the acquisition of the essential lexis to learning.

2. Sampling

To test lexis, the ATM proposes that a list of all important and frequent lexical-items that play a role in the understanding of English language, and that were learnt in lessons has to be elaborated. To this list, the new words that students have encountered while developing the language skills should be sampled. All this vocabulary stock would be a bank from which the test developer has to select which items to be incorporated in vocabulary tasks.

3. Possible methods

In the ATM, major test methods assessing vocabulary include examples as:

- Matching pairs (definitions, synonyms, opposites)

- Dividing words into their roots and other affixes.

- Classifying words according to alphabetical order.

- Crossing odd words in a given list.

- Adding more words to a list.

4.5.2.4 Test Format

4.5.2.4.1 Structure of the Test

The ATM is organized around three separate but complementary tests: **SUB-TEST ONE**, **SUB-TEST TWO**, and **SUB-TEST THREE**. SUB-TEST ONE is testing listening; SUB-TEST TWO is to test speaking. SUB-TEST THREE deals with testing the other skills (reading and writing), and other language components (grammar and vocabulary). (See Table 4.2).

A FRAMEWORK FOR ALTERNATIVE TESTING MODEL					
SUB-TEST ONE (Testing Speaking)	**SUB-TEST TWO** (Testing Listening)		**SUB-TEST THREE** (Testing Reading + Writing)		
Section One	**Section One**	**Section Two**	**Section One**	**Section Two**	**Section Three**
Listening Comprehension	**Speaking**	**Written Oral**	**Reading comprehension**	**Language Competence**	**Writing**
- Global comprehension - Detailed comprehension - Summarizing	- Factually-oriented speaking - Evaluative-oriented speaking	- Pronunciation **-** Rhyme - Syllables - Silent letters	- Global comprehension - Detailed comprehension - Summarizing	- Lexis - Morphology - Syntax - Discourse	- Mechanics - Composition

Table 4.2: The ATM: Structure

225

1. Sub-Test One

The structure of Sub-Test One (Testing Listening) consists of one section and one test. The gist of testing listening is to evaluate the test takers' ability to comprehend any kind of discourse, to discriminate phonemes, and to retain what has been heard. In other words, this means it is to judge the test takers' abilities to:

- Identify
- Select
- Judge
- Explain
- Synthesize

Sub-Test One has two parts: listening comprehension and listening accuracy. With listening comprehension, test takers are prompted to listen to a recording about a topic proposed in advance and with a great care, and asked to respond to the tasks. The items in this first part of the Sub-Test One are assessing both global and detailed comprehension. The aim is to see whether or not the test takers get the general and other secondary ideas of the topic. In the second part of this sub-test, the degree of accuracy of listening is checked. This part is carried out through two basic tasks that are dictation and summarizing.

2. Sub-Test Two

Sub-test Two (Testing Speaking) is the second test in the ATM. The basic aim of this test is to meet the requirements of the communicative needs set out in the relevant syllabus. It is important to develop tasks that more accurately reflect those aims. Also, it is crucial to note that throughout the design of Sub-Test Two consideration obviously has to be given to the learning context in which actual instruction takes part. The arrangement of the test allows for four-test takers to interact. This format into four-test takers group aims to get some advantages. For instance, it is easier to monitor the time allocated for the test. Besides, it helps to create a genuine information gap necessary for communicative interaction. There is

greater equality between test takers, which not only makes them more willing to interact, but also makes the communication as fluent and as successful as possible. As a result, this format arguably encourages a more natural interaction. The role of the teacher in this context is as an assessor. In some cases, s/he can be assisted by another fellow colleague for the purpose of marking. This assistance can make the test more reliable. But, in case this is not possible, one assessor (teacher) can suffice. (See Figure 4.1).
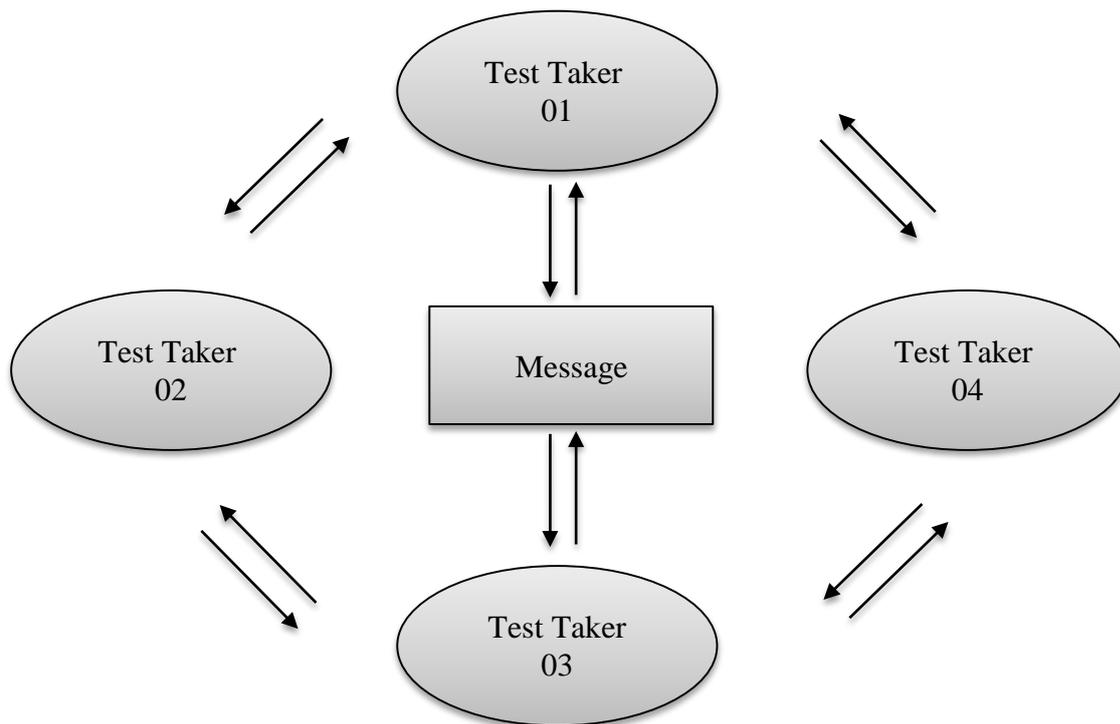
**Figure 4.1: The Pattern of interaction for the proposed
Sub-Test Two (Testing Speaking)**

The structure of Sub-Test Two is made-up of two separate sections. What is remarkable on this last point is that each section is held separately from the other one. But, the two make one. The sections are outlined below:

a. Section One

This section is formed of two dependent parts:

    i. Part one is about factually-oriented speaking

    ii. Part two is about evaluative-oriented speaking

In part one, each one of the four test takers briefly introduces himself/herself to the assessor, who then interacts with each test taker by asking further questions around a topic proposed in advance and taken from the programme taught in class. Thus, during this first part of Sub-Test Two (Testing Speaking), assessment will focus more upon individual or pair performance.

In part two, the purpose of the Sub-Test Two is to meet the communicative requirements through problem solving. The procedure adopted by the assessor is to make the test takers negotiate the solution to the problem. It will be necessary for them to alternate between transactional, long and short turns of speech. What is crucial in this part of the test is to set up information gap interaction tasks, and real-life tasks with the intention to have a favourable washback effect upon learning.

b. Section Two

The second section in Sub-Test Two is labelled 'Written Oral'. As stated earlier, this section is held at a different time from section One. It is mainly focusing on assessing discrete listening in a written production. The aim of this section is to evaluate the test takers' ability to recognize sounds, stress, and silent letters through the use of short items such as rhyme, syllable, and intonation.

3. Sub-Test Three

Sub-Test Three is about the assessment of reading, writing and other language components, such as grammar and lexis. This test comprises three sections that are:

- Section One: Reading Comprehension
- Section Two: Language Competence
- Section Three: Writing

a. Section One: Reading Comprehension

This section is directly related to a supporting test. It is divided into three parts:

    i. Part One concerns global comprehension

    ii. Part Two is about testing detailed comprehension

    iii. Part Three deals with the ability to synthesise through the summarizing of the text.

The basic purposes of this first section are to evaluate the tests takers' ability to:

- Identify
- Explain
- Illustrate
- Select
- Argue
- Judge
- Synthesize

b. Section Two: Language Competence

The second section is labelled 'language competence'. It is intended for the evaluation of some language components, such as grammar and lexis. It is formed of four parts:

    i. Part One concerns testing lexis

    ii. Part Two is about testing word formation

    iii. Part Three deals with testing sentence arrangement

    iv. Part Four tackles testing coherence and cohesive devices.

The basic purposes of this second section is to evaluate the test takers' ability to:

- Restate
- Recall
- Reproduce
- Predict
- Identify
- Construct
- Use
- Organize

c. Section Three: Writing

The last section in Sub-Test Three is about writing. It is related to the topic of the supporting text since this test is expected to follow a thematic approach. It is formed of two parts:

i. Part One concerns testing the ability to handle some writing clues.

ii. Part Two is about testing the ability to produce a comprehensive composition.

The basic purposes of this third section is to evaluate the test takers' ability to:

- State
- Respond
- Predict
- Use
- Construct
- Organize

- Argue
- Summarize
- Judge
- Support
- Attack
- Evaluate

What should be noted about the three sections in Sub-Test Three is that Sections One and Three are inherently dependent, whereas Section Two can be separate from the two other ones (see Appendix 7).

4.5.2.4.1.1 Task Distribution

In the ATM, the number of tasks in the three sub-tests is estimated to be 25. In sum, they are distributed as follows:

1 Sub-Test One: 8 tasks

2 Sub-Test Two: 4 tasks

3 Sub-Test Three: 13 tasks

A detailed presentation on the number of tasks per section is displayed in Table 4.3.

SUB-TEST ONE (Testing Listening)	SUB-TEST TWO (Testing Speaking)		SUB-TEST THREE (Testing Reading + Writing)		
Section One	Section One	Section Two	Section One	Section Two	Section Three
Listening Comprehension	Speaking	Written Oral	Reading comprehension	Language Competence	Writing
1. Listening comprehension a. Global comprehension b. Detailed comprehension **(Two tasks)** 2. Listening accuracy a. dictation b. summarizing **(Two tasks)**	1. Factually-oriented speaking a. Pair-work **(Two tasks)** 2. Evaluative-oriented speaking a. Interaction tasks b. Real-life Tasks **(Two tasks)**	1. Pronunciation **(One task)** 2. Rhyme **(One task)** 3. Syllables **(One task)** 4. Silent letters **(One task)**	1. Global comprehension **(Two tasks)** 2. Detailed comprehension **(Two tasks)** 3. Summarizing **(One task)**	1. Lexis **(Two tasks)** 2. Morphology **(One task)** 3. Syntax **(Two tasks)** 4. Discourse **(One task)**	1. Mechanics **(One task)** 2. Composition **(0ne task)**
4 tasks	**4 tasks**	**4 tasks**	**5 tasks**	**6 tasks**	**2 tasks**
4 tasks	**8 tasks**		**13 tasks**		

Table 4.3: The ATM: Task Distribution.

4.5.2.4.1.2 Time Allotted for the Test

The time estimated for the three sub-tests is about 5 hours: 1 hour for Sub-Test One, 2 hours for Sub-test Two, and 2 hours for Sub-Test Three. What is noticeable on this time allocation is that the three tests are held separately in different periods of time, and even in different days. For a clear picture of the time allotted to the ATM, see Table 4.4 below.

SUB-TEST ONE (Testing Listening)	SUB-TEST TWO (Testing Speaking)	SUB-TEST THREE (Testing Reading + Writing)
1 hour	2 hours	2 hours
5 hours		

Table 4.4: The ATM: Time allocation

The time estimated that Sub-Test Two takes is 1 hour for Section One (Speaking) , and 1 hour for Section Two (Written Oral). If things are clear for Section Two, for Section One, the ATM proposes a test estimated to take between 10 to 12 minutes for each assessed group of test takers, bearing in mind that each group is formed of 4 students and the allocated time here is applicable for a class of 20 students. In this context, it is likely argued that much reliable information in an oral test can be obtained in less than 15 minutes (Hughes, 1989: 15). A short oral test, such as the one proposed, is easy to administer as assessment could only take place during regular timetable class.

4.5.2.4.2 The Scoring Method

In the ATM, to maximize the usefulness of the test, one correct answer is recognized. This scoring technique is based on the approach that considers the operation of marking using a rating scale. The assumption that this approach underlies is to specify the criteria used for correction or the criteria by which a given response is to be assessed. For instance, the criteria for testing listening are not the same as those employed to test writing. A correct answer for a

listening item should take into account the comprehension aspect as a basic criterion to mark it, whereas in a composition an answer should be qualified on the ground of some criteria, such as grammar, vocabulary, coherence, and style.

1 Score Aggregation

As mentioned earlier, the ATM is made-up of three sub-tests. The weighing of the whole test will vary from one to another. This is so due to the fact that, for instance, sub-test one is about listening, whereas the third test encompasses more than one skill i.e., reading and writing. Besides, the time allotted for a listening test is shorter that the time set to sub-test three. Also, the number of tasks included in sub-test three is bigger than the two other tests. In this respect, the Alternative testing Model proposes to allocate different percentages to the three sub-tests. These are as follows:

Sub-Test One 25%, Sub-Test Two 25%; and Sub-Test Three 50% from the whole mark of the test. The final mark is an aggregation of these tests.

2 Setting Scores

In the ATM, the test is scored objectively by means of a key answer. The latter consists in eliciting the answer that corresponds to each test-item. A scoring scale is used to test the performances of the test takers in the three tests. Each sub-test is scored independently. The final scores of the three sub tests are calculated on the basis of their correspondent percentages; at the end of this operation, the final score is the sum of the scores obtained in each one of the three sub-tests. This way of conceiving the scoring method and scale aims at weighing exactly the performance of test takers in that foreign language, so that this final score would reflect their actual level. For an illustrative description of the scoring method proposed in the ATM, see Table 4.5.

SUB-TEST ONE (Testing Listening)	SUB-TEST TWO (Testing Speaking)		SUB-TEST THREE (Testing Reading + Writing)		
Section One	Section One	Section Two	Section One	Section Two	Section Three
Listening Comprehension	Speaking	Written Oral	Reading comprehension	Language Competence	Writing
1. Global comprehension (8pts) 2. Detailed comprehension (8pts) 3. Summarizing (4pts)	1. Factuality-oriented speaking (5pts.) 2. Evaluative-oriented speaking (5pts.)	1. Pronunciation (4 pts) 2. Rhyme (2 pts) 3. Syllables (2 pts) 4. Silent letters (2 pts)	1. Global comprehension (2pts) 2. Detailed comprehension (2pts) 3. Summarizing (2pts)	1. Lexis (2pts) 2. Morphology (2pts) 3. Syntax (2pts) 4. Discourse (2pts)	1. Mechanics (2pts) 2. Composition (4pts)
20pts.	10pts	10pts	06pts	08pts	06pts
20/20	20/20		20/20		
25%	25%		50%		

Table 4.5: The ATM: Scoring scale

Conclusion

In sum, this chapter offered a theoretical background of the Alternative Testing Model. It brought an understanding of how the art of developing useful language tests is very systematic. It addressed key issues that any test developer should consider to make his/her tests ready for application in order to draw out the intended inferences. It displayed major components and factors that are needed in the process of test development in order to avoid any anomalies that might affect the teaching and learning processes. In short, what could be retained from this chapter is that the intention to propose an ATM in this study allowed a truly fair and informative check of the learners' attainment and progress in English language in accordance to what these learners have dealt with in their classes, and also in relation to the relevant instructional objectives set out in the used syllabi.

Now that the ATM is made clear enough for its users, the next phase in this study will be to submit it for trial in order to judge more or less its usefulness for the intended purposes. Hence, the following step will tell us more about the kind of washback effect that this new testing system would exert on teaching and learning in the context under investigation.

Chapter Five: Teachers' and Students' Reactions to the Alternative Testing Model Implementation

Introduction

This chapter reports the results on the teachers' and students' reactions to the Alternative Testing Model implementation, and discusses and summarizes the findings. Unlike to the structure of chapter 3, reporting the findings in Phase One, in this chapter, in addition to questionnaires and classroom observation, interviews and focus groups are employed as data collection methods. Chapter 5 comprises four sections. Section one elicits the research questions and purposes that shaped and limited the scope of study in this phase. Next, section two describes the research design of the final study, including a presentation of the design, structure, validation, and analysis procedures of the adopted instruments. Following this section, section three reports the results from all the data collection methods. Finally, the chapter ends with a discussion and summary of these findings.

5.1 Research Questions and Purposes of Phase Three

Phase Three in this study began after Phase One and Two had been completed. The research in this phase explored teachers' and students' reactions to the ATM implementation. In precise terms, it investigated whether or not the implementation of the new policy had any washback effects on classroom teaching and learning in terms of teachers' and students' behaviours. The research questions were formulated as follows:

> RQ3: What is the nature and scope of the washback effect on teachers' behaviours as a result of the ATM implementation?

> RQ4: What is the nature and the scope of the washback effect on students' behaviours as a result of the ATM implementation?

Phase Three -also referred to as the Final Study- here had the practical aims of gathering evidence in order to:

- identify the aspect of teachers' behaviours in the classroom as a reaction to implementing the ATM. These included:

- teachers' reactions to, and perceptions of, the new testing system;
- teachers' materials;
- assessment and evaluation; and
- teachers' attitudes towards aspects of learning.

- display students' reactions towards, perceptions of, the ATM. These include:

- students' learning contexts;
- students' attitudes towards teaching and learning, inside and outside class;
- students' attitudes towards aspects of learning; and
- students' perceptions of achievement tests based on the new testing model.

5.2 Research Design for the Final Study

5.2.1 Data Collection Instruments

5.2.1.1 The Questionnaires

5.2.1.1.1 Rationale and Aim

Teacher's Questionnaire (02) (TQ$_2$)

The purpose of the teacher's questionnaire (02) (TQ$_2$) is to explore teachers' perceptions of, attitudes towards, and reactions to, the implementation of the ATM. In precise terms, this questionnaire could be an appropriate method in order to allow a considerable number of teachers in the context under exploration to voice their opinions on the new testing policy.

More importantly, the adoption of this data collection method seeks to investigate whether or not any washback effect of the proposed testing system has occurred on teachers perceptions of aspects of classroom teaching.

Student's Questionnaire (02) (SQ$_2$)

The purpose of the student's questionnaire (02) (SQ$_2$) is to investigate the students' perceptions of, attitudes towards, and reactions to, the implementation of ATM. In precise terms, the student's questionnaire would be an appropriate method to allow a considerable number of students in the context under study to voice their opinions on the new suggested testing model. Crucially, the employment of this data collection method seeks to identify whether or not any washback effect of the ATM has occurred on students' perception of aspects of classroom learning.

5.2.1.1.2 Structure and Contents

Teacher's Questionnaire (02) (TQ$_2$)

The teacher's questionnaire (TQ2) for the main study consisted of four parts and was designed in the target language. Section One (TQ2-1-7) consisted of 7 questions and 21 items about background information covering (a) demographic details, such as gender, age, academic qualifications, teaching experience, (b) current teaching situation, such as levels taught, teaching sessions per week, and teaching hours per week. Section Two (TQ2-8-9) consisted of 2 questions and 21 items. The two questions employed Likert scale. The latter is a useful device for the research, as it builds in a degree of sensitivity and differentiation whilst still generating numbers. Also, this way of asking questions is efficient since it combines the opportunity for a flexible response with the ability to determine frequencies, and other forms of quantitative analysis. It affords the researcher the freedom to fuse measurement with opinion, quality with quantity (Cohen et al., 2005:253). Using a four-point

scale in this study rather than a traditional five-point or seven-point scale is a way of controlling for the 'central tendency', a tendency for participants with no strong feelings to opt for the mid-point of a scale. This has been found to be a particular problem with many respondents, where the 'mean' is advocated, so there are tendencies to respond on the basis of the form rather than the content of items. This part of the questionnaire aimed at exploring the teachers' perceptions of the new testing policy.

Section Three (TQ2-10-11) consisted of 2 questions and comprised 14 items; the two questions were designed on a four-point Likert scale. Teachers were invited to choose the appropriate answer according to their own teaching situations. This part of the questionnaire aimed at investigating the washback effect of the ATM on teaching materials.

Section Four (TQ2-12-14) consisted of 3 questions and comprised 10 items. One question (TQ2-12) was designed using a multiple-choice format, whereas, the two remaining questions (TQ2-13-14) were designed on a four-point Likert scale. This part of the questionnaire aimed at exploring washback on classroom teaching behaviour.

For clarity and simplicity of reporting, notation as (TQ2) is used to describe teachers' questionnaire in Phase Three - the Final Study- and (TQ2-2) is used to describe teachers' questionnaire in Phase two, Question 2. The structure and themes of the teacher's questionnaire (02) (TQ2) are summarized and illustrated in Table 5.1 below:

Structure	Contents
Theme One	**- Personal Particulars (TQ2-1-7).** • Demographic information: gender, age, academic qualifications, teaching experience. • Current teaching situation: English level currently taught, the number of sessions taught per week, and teaching hours per week.
Theme Two	**- Teachers' Perceptions of the New Testing System (TQ2-8-9)** • Learning aims. • Learning strategies
Theme Three	**- Teaching Materials (TQ2-10-11)** • Teaching materials in terms of resources • Learning materials in terms of resources
Theme Four	**- Teachers' Classroom Behaviours (TQ2-12-14)** • Medium of instruction • Teacher talk • Teaching activities

Table 5.1: The Final Study structure and themes of the teacher's questionnaire (TQ2)

Questionnaire for Students (02) (SQ2)

The student questionnaire (SQ2) for the Final Study consisted of two parts and was designed in the native language of the respondents, Arabic, but translated into English for further analysis. Section One (SQ2-1-5) consisted of 5 questions and 14 items about students' demographic information and their learning context. Students' demographic information included the gender, the student's level, the branch where the student studies in, the number of hours a student undertakes to study English at school per week, and the number of hours a student spends for revising at home.

Section Two (SQ2-6-12) consisted of 7 questions and 37 items; the questions employed a four-point Likert scale. This part of the questionnaire aimed at exploring the students' attitudes towards teaching and learning activities inside and outside the classroom in relation to the ATM.

For clarity and simplicity of reporting, SQ2-2 is used here to student's questionnaire in the Final Study, Question 2. The structure and themes of the student's questionnaire (02) (SQ2) are summarized and illustrated in Table 5.2 below:

Structure	Contents
Theme One	**- Demographic Information and Learning Contexts (SQ2-1-5)** • Demographic information: gender, grade, branch. • Current learning situation: number of hours students undertake to study English at school per week, and number of hours students spend for revision.
Theme Two	**- Students' Attitudes Towards Aspects of Learning and Teaching Activities Inside and Outside the Classroom (SQ2-6-12)** • Teachers' talk in class • Students' perceptions of their teachers' teaching activities in class • Students' use of English in relation to their classroom activities • Students' attitudes towards aspects of learning • Students' preferred learning strategies • Students attitudes towards English tests

Table 5.2: The Final Study structure and themes of the student's questionnaire (SQ2)

5.2.1.1.3 Piloting and Validation of the Instrument

The pilot study was carried out in June 2012. The aim of this crucial stage in a questionnaire development was to ensure validity and reduce problems with the procedures of administering the questionnaire. In addition, another aim was to achieve a focus on the research questions and the relevant research purposes. In doing so, the researcher also sought to determine (a) the various items in the questionnaire were understood as intended by the researcher, and (b) they were understood in the same way by each teacher and student. This technique could enable the researcher to check out whether the contents and layout are clear

241

for the respondents, and to determine what was and was not working in this data collection method.

In this respect, concerning the teacher questionnaire, the questionnaire was submitted to two university experts. Both of them hold a PhD degree and had a considerable experience as supervisors and university teachers. One expert was from the University of Ouargla (Algeria), and the second from the University of Adrar (Algeria). It should be pointed out that these two university teachers did not participate in the piloting stage of any data collection instruments for the Preliminary Study. Besides, the teacher's questionnaire was also handed to two fellow doctoral researchers from the University of Biskra (Algeria), and 10 EFL teachers from the secondary schools in Biskra region. In sum, 14 people participated in expertizing and reviewing the teachers' questionnaire (TQ2) for the Final Study. The participants were given enough time to read and comment. The methods employed to submit the questionnaires for the experts consisted of handing them directly or sending them via emails. For convenience with the ethical guidelines followed throughout this research, which posit that the identity of people who take part in the piloting stage should be kept anonymous, the researcher replaced the names of the experts with correspondent pseudonyms, following the order that started in Phase One i.e., E_7, E_8, E_9, and E_{10}. Table 5.3 below illustrates the participants, their degrees, and the place where they work.

No	Participants	Degree	Place of work
01	E_7	PhD	University of Ouargla
02	E_8	PhD	University of Adrar
93	E_9	Magister	University of Biskra
04	E_{10}	Magister	University of Biskra
05	10 EFL Teachers	Licence (BA)	Secondary schools in Biskra region

Table 5.3: Participants in the pilot study for the teacher's questionnaire (TQ2)

In terms of remarks and comments of the experts and the participants in the piloting stage concerning the content and layout of the teacher questionnaire (TQ2), these were as follows:

E_7 found that the terminology and language used in the teacher's questionnaire was clear, and did not pose any problems to respondents. However, in terms of the questionnaire's length, E_7 thought that the latter is a little lengthy to respond. He advanced the idea that the more a questionnaire is long, the more it becomes boring for the respondent to answer. Because of this remark, he thought it would be more appropriate to review the number of questions and items and try to fuse the ones that bore the same sense.

E_8 converged to a large extent with E_7's remark, but in addition to what E_7 said, E_8 was more precise in eliciting a very significant remark. He thought some questions were not correspondent to the nature of the analysis that the researcher intended to carry out. For instance, one of the questions in Section One of the questionnaire was designed in an open-ended format. Bearing in mind that this type of questions do not correspond to a quantitative description, it would be difficult for the researcher to adapt it to the requirement of descriptive statistics, as the researcher intended to do it further in the analysis phase. Due to this remark, E_8 suggested to re-word this question and make it resemble to a multiple-choice format, or a four-point Likert scale, or simply omit it. After reading carefully this remark, the researcher decided to omit this question from Section One and limit himself only to close-ended questions.

E_9 did not object any of the questions. The only remark he did was that the researcher had to be certain that the sectioning of the themes should be in total congruence with the research purposes. Contrary to E_9, E_{10} was not explicit in his evaluation of the teacher's questionnaire. This doctoral researcher expressed some doubts about some of the terminology

in English. He asserted that it might be difficult for some respondents to understand the right senses of terms, such as 'washback' or 'ATM'. He added that it would be better if the researcher drew lines in the items where the respondent was expected to give details. This would give the respondent more freedom to say whatever s/he thought was the most important. For the researcher, it was not possible to leave spaces for justifications or explanations because of the data he sought to obtain, and the nature of the analysis he intended to carry out which required only quantitative final outcomes.

Finally, for the 10 EFL teachers in the secondary schools in Biskra region, they did not express any doubts or objections. As a result, the content and format of the teacher's questionnaire (TQ2) was kept as it was. The only changes were mentioned previously with E_7, E_8, and E_9. For a final copy of teacher's questionnaire (02) in the Final Study, see Appendix 8. The design stages and validation procedures for teacher's questionnaire (TQ2) are summarized and illustrated in Figure 5.1 below:
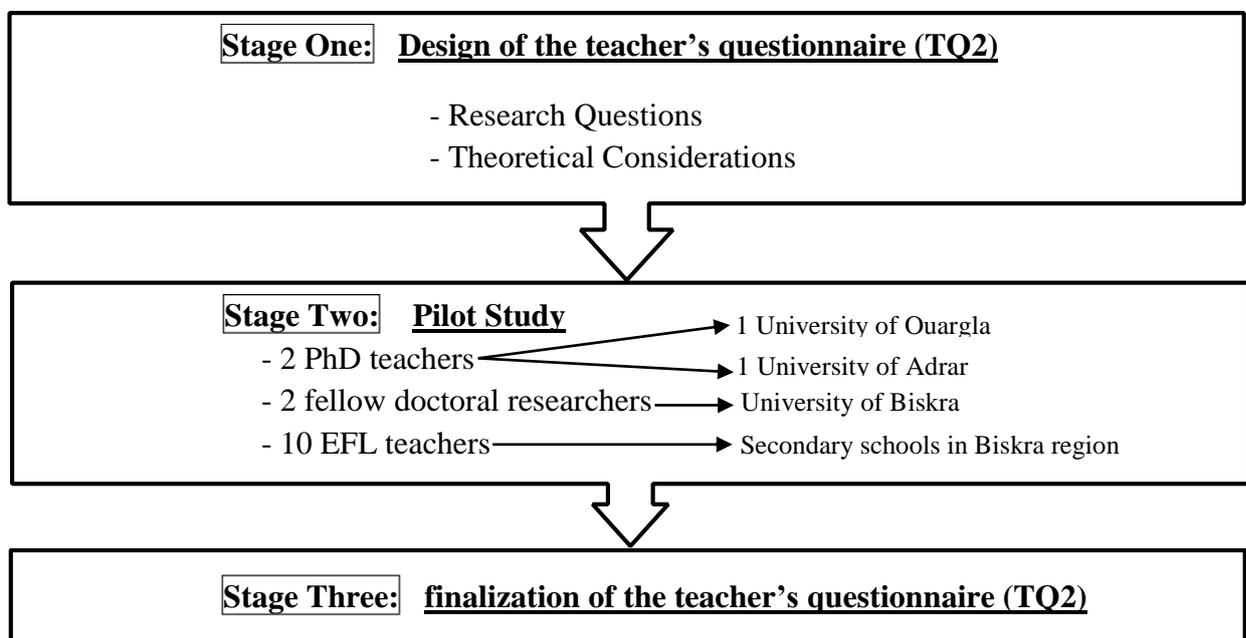
Figure 5.1: Stages of the teacher's questionnaire design for the Final Study.

The student's questionnaire (SQ2) for the pilot study was designed parallel to the teacher's questionnaire. Therefore, the basic considerations for validation were the same as for the teacher's questionnaire. The piloting stage was carried out rather differently from that in the teacher's questionnaire, however, as the student's questionnaire was administered in Arabic rather than English. The rationale behind this was to minimize misunderstanding and reliability and validity problems caused by the language factors. The purpose of the validation and piloting phases was to try-out the questionnaire by attempting to answer the questions and provide comments on any problems the respondents had and any issues that occurred to them as they did so. Besides, the other aim was to pre-test the questions and items, along with ease of use, format and overall instrument appearance, and to determine what was and was not working in the questionnaire. Above all, this procedure served as a method of construct validation for the questionnaire items and to evaluate specific question wordings.

The 13 people involved at this stage were 1 general inspector of English language, 2 doctoral candidates, and 13 EFL students from the target population, but not part of the final sample. All the participants were invited to complete the questionnaire and voice their opinions. Enough time was given to them to read and note potential problems and also identify the specific points where they might have difficulties and the type of misinterpretation one might make. The students were asked to try-out the questionnaire with the presence of their teachers. The purpose was to enable them to ask questions about the questionnaire or the task while they were doing it. The questionnaires were handed directly to them, and given back to the teachers when they were filled in by the students. The same means was used with the other participants. The only difference was that the questionnaires submitted to the general inspector and the 2 doctoral candidates were original copies, not the

translated forms. Table 5.4 illustrates the participants, their degrees, and the place where they work.

No	Participants	Degree	Place of work
01	E_{11}	General Inspector	Batna region
02	E_{12}	Magister	University of Biskra
03	E_{13}	Magister	University of Biskra
04	10 EFL Students		Secondary schools in Biskra region

Table 5.4: Participants in the pilot study for the student's questionnaire (SQ2)

The general inspector involved at this stage expressed doubts about the length of the questionnaire. He expressed impatience and lost concentration in filling in the questionnaire. He added that questions in Section Four would be better asked in interviews to gain positive feedback, rather to be included in a questionnaire where it is expected that respondents could write detailed answers. As a result, a change in the content of the preliminary version in the pilot stage took part. Section Four, comprising 8 open-ended questions were omitted. E_{12} remarked that the idea in Section Two in the student's questionnaire did not match with the general purpose of this section in that Section Two aimed to explore students´ attitudes towards teaching and learning behaviours, whereas question 8 dealt with resources in class. Therefore, this question was removed from the questionnaire. In addition, the same expert noted that SQ2-13 in Section Two was too vague and hence confusing to respondents. He suggested either a rewording or a complete omission of this item. Due to this remark, the researcher thought that such a question would be better reworded and removed to be fused with SQ2-13 in Section Three.

E_{13} mentioned because of redundancy between SQ2-15 and SQ2-17, it would be better to fuse these two items in one item or simply drop out the first one. As a result, the researcher decided to leave off SQ2-15 and keepSQ2-17 because the latter sounded more comprehensive than the left item. Usually with E_{13}, this expert noted there were also redundant areas in three

items in Section Three. These were SQ2-25, SQ2-26, and SQ2-27. Due to this remark, these questions were also left off the questionnaire. 10 EFL students did not object answering the questions; and they did not provide any comments or remarks on the content and format of the questionnaire. They found it easy to read and understand the questions since the questionnaire was translated for them into their native language, Arabic. The final version of the student's questionnaire (02) in English included 12 questions as shown in Appendix 9. Figure 5.2 illustrates the stages of the questionnaire design together with the validation procedures discussed above.
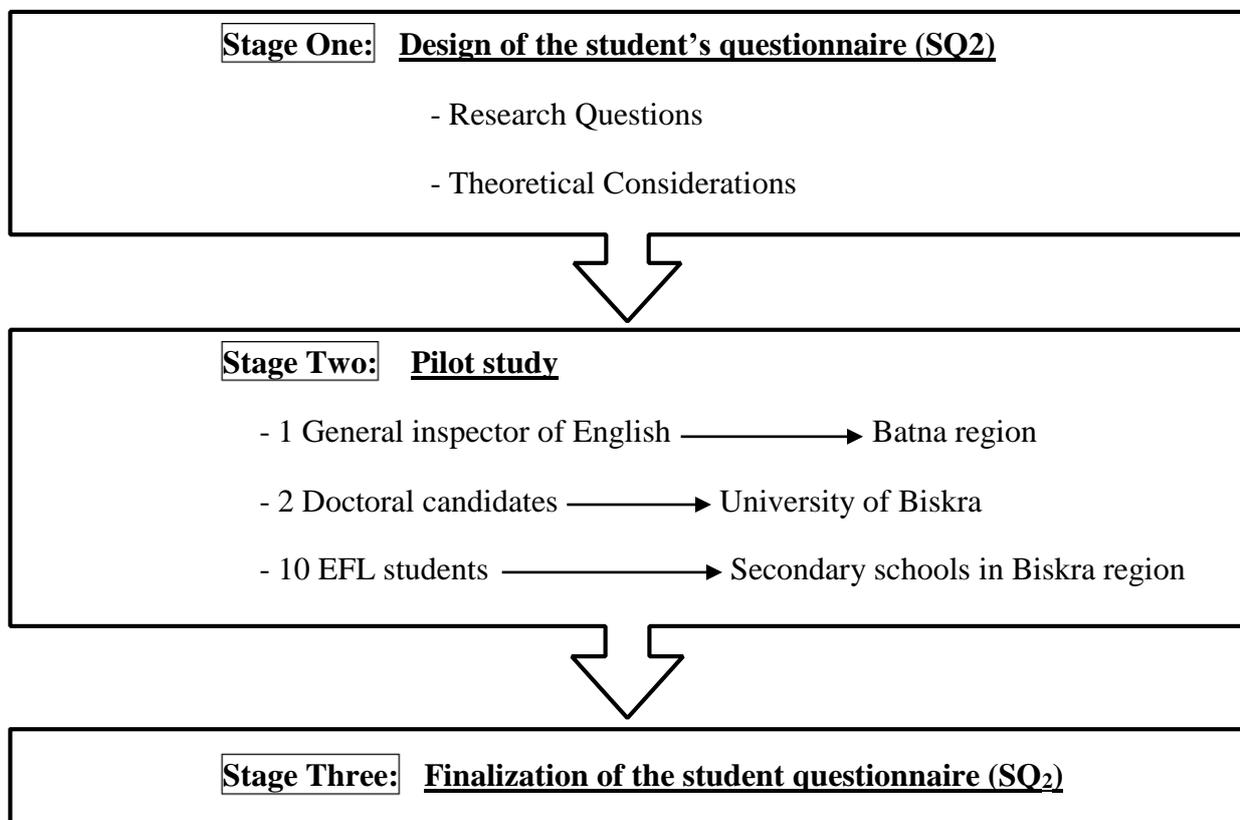
Figure 5.2: Stages of the student's questionnaire design for the Final study

5.2.1.2 The Interviews

5.2.1.2.1 Rationale and Aim

Interview with Teachers (02)

The general aim of the interviews in the main study was to investigate the breadth and range of views represented by the participants of the complex topic of washback in relation to the implementation of the ATM. With teachers´ interviews, the purpose was to explore the teachers´ beliefs: whether teachers believed that their teaching had been influenced by the new testing policy? and whether they had made changes in their teaching practices as a result of the new model? Besides, these interviews also provided for the teachers an opportunity to give their impressions on the content of what they taught, and to express their opinions regarding the development of achievement tests for their students based on the ATM.

Interview with Students (02)

In parallel to the interviews with teachers, the interviews with students in the Final Study were semi-structured. They were conducted in the target language. The participants were allowed at some occasions to use their native language. The purpose of the interviews with students was also to explore the students´ beliefs: whether students believed that their learning had been influenced by the introduction of the ATM, and whether they had made changes in their learning practices as a result of the new testing policy? The interviews also provided for students were an opportunity to give their impressions on the lessons through which they are constantly taught, and to describe the activities and materials their teachers adopted regarding the intervention. Besides, the interviewees were invited to voice their opinions about the difference that they might have noticed between the current testing system and the ATM. Finally, the interviewed students were asked to yield their views on whether or not to adopt this model.

5.2.1.2.2 Structure and Content

All of the interview questions were from the review of literature, and the preliminary-gathering stage. There were parallels between questions in the questionnaires and interviews. In precise terms, the content of the questionnaires guided the researcher to design the two interviews in this phase. The purpose was to corroborate what was obtained as data in the questionnaires, as well as to complete what was thought to be lacking or absent in these data, and propose questions in the interviews to enable respondents to provide these information. In this respect, the structure of each of the two interviews with teachers and students were determined.

Interview with teachers (02)

As mentioned previously, the interviews with teachers in the Final Study were of semi-structured type. This type of interviews is often regarded to be suitable when the researcher has a good enough overview of the phenomenon or domain in question and is able to develop broad questions about the subject under study in advance, but does not want to use ready-made responses categories that would limit the depth and breadth of the respondents´ story (Dörnyei, 2007:136). Because the main purpose of the interviews with teachers in Phase Three of this study was to explore the participants´ views on the complex topic of washback due to the introduction of the new testing policy, the quasi-totality of the proposed questions were designed to enable these participants to generate their opinions. Thus, the list of these questions included items that started with 'wh-questions', such as 'how', 'why', or used introductory phrases, such as 'in your point of view', 'in your opinion'. Besides, in order to offer the interviewee the chance to relax, the interviewer thought to begin the interview with an easy question. That´s why, in the interview schedule, the first question was about something the interviewee might be expected to have well-formulated items and near to his/her knowledge.

Interview with Students (02)

Having decided on the interview type i.e., semi-structured interviews, with students during this phase of study, the researcher started drawing-up a list of questions based on the research questions identified earlier. At a theoretical level, this process is known as "the interview guide approach" (Patton, 1987). In practical terms, this process is usually used to talk about and discuss; it does not provide the order of either questions or wording. For the present interview, the interview guide comprised 14 questions. These questions constituted the basis on which the students were given the opportunity to express their views and opinion on the previously identified issues. Overall, three themes were expected to be discussed. These were: Theme One, concerned with the students´ perceptions of the new testing model; Theme Two dealt with aspects related to test content and format; and Theme Three sought to voice the students´ impressions on the influences the ATM might have exerted on learning.

5.2.1.2.3 Piloting and Validation of the Instrument

Before embarking on the pilot study, the researcher sought to obtain an oral consent from the participants. With no hesitation, the interviewed students showed readiness and devotion to take part in the interviews. For the researcher, things were easier than he expected them to be. This was due to some reasons. First, the researcher felt he was involved in the scene and with the participants, as if he were one member of the group. Second, because the purposes and procedures of the research became clear, the participants were motivated to contribute in this study.

Practically, the next step was to decide who should be participants in the pilot study. To respond to this question, the researcher opted to choose 2 secondary fellow teachers and 5 students from each of the classes that were part from the targeted population. All these formed the sample of the interviewees who were expected to respond to the questions of the

two interviews in order to make them ready for final use. At this stage of the interview development, the participants in the pilot stage were asked to respond the questions and to comment on any problems they had, as well as on any issues that occurred to them after the interviews. Besides, for the purpose of examining the reliability of the interview data, the researcher went back to the audio-recordings, and checked again the transcripts. The purpose of this procedure was to see whether or not he had been consistent with the criteria of interviewing and analysis. Agreement was reached in about 99% of the contents in both interviews.

In terms of results, the two teachers involved in this stage expressed doubts about the number of questions included in the interview. They felt that it took too much to answer these questions. As a result, the content of the interview for teacher was altered. For instance, Q6, Q11, Q15, Q18, and Q19 were removed. The aim of this reviewing was to make the content of the interview match with general purpose of this data collection method, and get rid of the questions that were repeated or those that overlapped with other questions. In sum, the final version of the interview for teachers in the Final Study comprised 12 questions. The final copy is presented in Appendix 10.

About the interview for students, 25 involved at this stage expressed no difficulties in answering any of the questions, while 10 indicated minor problems with wording. The question was then adjusted and a number of substantial adjustments and modifications were made to the interview items. The students also mentioned that they felt that the number of questions was big that they could not answer them thoroughly. Due to this remark, and after a meticulous reviewing of the questions, the researcher opted to leave off some questions. These were as follows. Q1, Q6, Q8, Q9, Q10, Q11, Q13, and Q14. In sum, the final version of the interview for students in the Final Study included 7 questions. The final copy is presented in Appendix 11.

251

5.2.1.3 The Observation

5.2.1.3.1 Rationale and Aim

This instrument sought to provide insightful feedback about classroom instruction in relation to the implementation of the ATM. Other specific purposes of this data collection method would be to see whether or not teachers assign more practice opportunities to students where those latter engage to communicate in English language, assign more class time to student activities, such as pair and group discussion which could increase language practice opportunities, assign more time to these activities to enable students improve their listening abilities and oral competence, and, finally to make students use more oriented and authentic materials to develop their communicative skills and capacities outside the classroom. In brief, classroom observation could be a good means to check out whether or not washback effect of the ATM on teaching and learning in the context under exploration has occurred.

5.2.1.3.2 Structure and Content

As discussed in Chapter 3, the adopted classroom observation is the same as the one described in 3.2.1.2.3. In the Preliminary Study of this research, Part A of the COLT was used since a classroom analysis at this stage matched the nature of the research questions in this phase to be answered. In precise terms, Part A of the COLT described classroom events at the level of activity. The categories based on this part were designed to (a) capture significant features of classroom events in EFL classes in the secondary schools in Biskra region, and (b) provides means of recording classroom interaction. The observation scheme in the Final Study consisted of five major categories: time, participant organization, activity type, content, and materials used. They were all coded in the classes. The researcher simply ticked under the category of participant organization and materials used during the

observations, but made some detailed field notes under the category of time, activity type, and activity content.

5.2.1.3.3 Piloting and Validation of the Instrument

The classroom observations for the Pilot Study started just after the introduction of the new testing system. As what was done in the Preliminary Study, permission to collect data was obtained orally at the level on individual teachers. The scheduling of each classroom observation was agreed in advance between the researcher and the teachers. All school visits were pre-arranged in with the teachers involved. Due to the difficulties to obtain permission from the school administration to video record lessons, only field notes were employed in its place for the pilot study.

One teacher, who was a member of the target population but not a part of the final sample, agreed to participate in the observation. 2 from the 8 classes were observed at this pilot stage. Each class was observed for 2 lessons, 60 minutes each. The purpose was to pilot the observation scheme and to gain familiarity with recording methods. At the end of the pilot stage, no problems were found in the procedures for using the observing or field notes recording methods. It should be pointed out that things got easier since this method of data collection had been employed since the beginning of the investigation in the Preliminary Study. As a result, the adopted observation scheme based on Part A of the COLT was kept for the Final Study without any changes.

To enhance the reliability of the data collection and analysis in the Final Study, the researcher returned to the field notes and checked them after 2 weeks later. The purpose was to make sure that the procedure of observing and analysis were carried out appropriately. Agreement was reached in about 95% of the classroom activities. Reaching this percentage proved that the researcher was consistent. In addition, reliability was enhanced because this

data collection method continued to be employed in 2 successive school years: 2010/2011 and 2011/2012. This made the researcher trained to collect data.

5.2.1.4 The Focus Groups

5.2.1.4.1 Rationale and Aim

For the purposes of the present study, it was felt that since the information was complex but involved a topic that was much already known to the participants, a data collection method using focus groups proposed in advance would be the best approach to culminate this phase. This instrument would allow additional, clear, and insightful data to emerge between the respondents but would also be flexible enough to allow informing or misinforming the collected information using the other employed data collection methods, such as questionnaires, observations, and individual interviews.

Since this method has such a pivotal role in eliciting rich information to make the discussion fruitful and is likely to draw a picture on the inferences that the implementation of the ATM had made on both teachers and students, the main purpose of focus group in the Final Study would be to display the respondents´ beliefs and their relationship with classroom practice. Specifically, for some practical reasons, participants would inevitably produce answers to examining their beliefs through a discussion of particular aspects of the topic under exploration. In this case, every respondent should interpret the ATM in his/her own way and make sound judgments on how useful is this innovation.

5.2.1.4.2 Structure and Content

To design focus groups for the Final Study, the present researcher followed some steps. These included the following:

- Step One: Determining the study purpose
- Step Two: Determining the participants

- Step Three: Preparing the setting

- Step Four: Obtaining informed consent

It should be pointed out that in the Final Study the respondents were selected only from the classes that tried out the new testing policy: each focus group included between 10 to 12 students. Though they were selected with opportunity and convenience taken into account, the researcher recruited participants with different levels and from various backgrounds. To do that, the researcher asked the help of the teachers since they knew better than other people the different academic, psychological, and social makes-up of their classes. In sum, the researcher opted to limit the number of groups to 2, for he thought that such a number was sufficient to gain information on the participants´ beliefs, attitudes, reactions, and practices in relation to the new testing policy.

5.2.1.4.3 Piloting and Validation of the Instrument

In order to make the focus groups schedule ready for final use, the researcher proceeded for the piloting of the first version. All the questions in this schedule were open-ended with prompts whenever necessary. All of them were conducted in English. The participants were asked to trial the questions by answering them and they were also invited to comment on any problems they had, as well as on any issues that occurred to them after the discussion. The purpose was to determine whether the questions were understood as intended. All the focus groups were audio-recorded and then transcribed by the researcher. The transcripts were edited by avoiding repetitions and putting substantive statements in chronological order to make grammatical sense, which helped the researcher in analysing the data further.

At this pilot stage, participants in focus groups expressed no difficulties in answering any of the questions. They claimed that they felt at ease when discussing the different topics raised by the questions, and they did not object answering any of the questions. Because of this focus group schedule was kept without any changes or modifications.

5.2.2 Respondents

Based on the nature and the methodological considerations of the present study, criteria to identify potential schools were selected:

- schools which comprised EFL classes;

- schools which agreed to accept to take part in this study;

- schools which agreed to accept that selected EFL classes could be observed;

- schools which agreed to accept that teachers and students would be allowed to participate in questionnaires;

- schools which agreed to accept that teachers and students would be allowed to participate in interviews and focus groups;

- schools which agreed to collaborate with the researcher; and

- school administrators who had a positive attitude towards the study.

Initially, 6 secondary schools were sampled from the 15 that included EFL classes. The 6 schools comprised 33% of the above schools population. This sampling served as a focus study. The strategy followed in choosing these schools took into account that these last were equally distributed in the different locations in Biskra region. From these 6 schools, 8 EFL classes were selected to take part in the investigation. All these classes were from the second year level. The rationale for this decision was that the other third year level was not appropriate for this study because at this level, classes would sit for the 'Bac' exam at the end of a school year. This would create great pressure on the participants that could not serve the purpose of this study.

What is notable about the teachers, who participated in this investigation, was that they were the same people who took part in the Preliminary Study. It was not possible to change them because there was a need to carry out the study with the same participants in order to check out any behaviour differences between the time that preceded the implementation of

the new policy, and the time that followed this intervention. The people who participated in this phase of the exploration were teachers with different backgrounds and teaching experiences. For convenience, a short description of the 8 teachers is going to be displayed below. It should be pointed out that the real identities of these teachers were kept anonymous, considering the agreement made with them at the beginning of the study, and following the ethical issues that guided this study. Pseudonyms were used to identify them.

- Teacher 1: T_1 is the youngest of the participants (23 years old) only having worked for 2 years. He holds a ´licence´ in English teaching. He is the only teacher who got a training to teaching before being hired. He graduated from the National Training School in Constantine (Algeria).

- Teacher 2: T_2 is 43 years old. He is the more experienced teacher among the participants. He has been teaching at the secondary school level for 18 years. He participated in many in-training-services, workshops, and seminars. He is the coordinator of the subject in his school. He holds a ´licence´ in English language from the University of Batna (Algeria).

- Teacher 3: T_3 is 31 years old. She has a good experience in teaching English. She has been teaching since 2005. She is the coordinator of the subject in her school. She holds a ´licence´ in English language from the University of Biskra (Algeria). In addition to teaching at the secondary school, she is a part-time teacher at the University of Biskra. She teaches the course of ESP at the Department of Foreign Languages.

- Teacher4: T_4 is 34 years old. He has been teaching for 3 years. He holds a ´licence´ in English. He graduated from the University of Biskra (Algeria).

- Teacher 5: T_5 is 29 years old. He has been teaching for 4 years. He holds a ´licence´ in English from the University of Biskra.

- Teacher 6: T_6 is 40 years old. She has a good experience in teaching. She has been teaching since 1999. She graduated from the University of Ouargla (Algeria).

- Teacher 7: T_7 is 35 years old. She has been teaching since 2009. She holds a ´licence´ in English from the University of Batna (Algeria).

- Teacher 8: T_8 is 29 years old. She has been teaching English for 3 years. She graduated from the University of Biskra.

In what follows, a more detailed description of the different respondents in data collection methods employed in this study is presented below:

5.2.2.1 Respondents to the Questionnaires

8 questionnaires were issued to teachers adopting the ATM. All of them completed the questionnaires, responded to the questions, and returned them back to the researcher, representing a return rate of 100%. Likewise, 300 questionnaires were issued to students who would take the ATM. The return rate was 58.33% (175 out of 300 questionnaires were completed and returned back to the researcher).

5.2.2.2 Respondents to the Observation

Similar to what was carried out in the Preliminary Study, 8 teachers and classes were observed. The number of sessions observed was estimated to 55 during the school year 2011/2012. In order to minimize the differences in observation to make the teaching more comparable, only 4 observed classes were used for the Final Study.

5.2.2.3 Respondents to the Interviews

All the 8 teachers who participated in the Final Study were interviewed. For the students, a randomly selected sample of 40 students were chosen. The sample consisted of

choosing randomly 5 students from each one of the 8 classes that took part in the investigation. The students selected were representative of different groups of the population.

5.2.2.4 Respondents to the Focus Groups

Because it was not possible to carry out focus groups with teachers due to some administrative constraints encountered by the researcher, this data collection method was only limited to students. Two focus groups were formed. Each group comprised 10 students. Overall, 20 students participated in this data collection method. The researcher operated as a moderator, and a fellow colleague as a notetaker. The choice of the participants was carried out randomly, considering that the two groups were representative of the population.

5.2.3 Data collection Procedures

5.2.3.1 Data Collection Procedures for the Questionnaires

Two different ways for submitting the questionnaires were used. The first way consisted of handing the questionnaires to the participants. They were given enough time to read and answer the questions. The second way was through emails. Teachers were asked to complete the questionnaires and email them back to the researcher. To ensure confidentiality and conform to the general ethical guidelines for conducting research, the participants were told in advance not to write down their names or display their identities.

5.2.3.2 Data Collection Procedures for the Interviews

The respondents were informed about the purpose of the interview. All the interviews were audio-recorded. The time devoted to one interview was between 20 to 25 minutes. After finishing the interviews, the researcher transcribed the audio-texts. For the sake of success of these interviews and compliance with the ethical guidelines for conducting research, the

respondents could object answering any of the questions. Their rights of confidentiality were protected.

5.2.3.3 Data collection Procedures for the Observation

In order to make the observation sessions well schemed, the researcher pre-arranged with the observed teachers the timetables for the observation. They were told in advance that the researcher would be present in the back of the classroom as a non-participant. He had not to intervene or comment while the teacher was presenting his/her lessons. The teachers could have a look at the notes the researcher took at any time they wanted.

5.2.3.4 Data Collection Procedures for the Focus Groups

The researcher explained to the participants what was expected from them to do. He told them that they had to feel relaxed and what was important was that they had to express their opinions and ideas at maximum, the researcher audio-recorded the focus groups. The note taker was there to help him in collecting the information. After finishing the focus groups, the researcher proceeded to transcribe the audio-texts.

5.2.4 Data Analysis procedures

5.2.4.1 Data Analysis Procedures for the Questionnaires

As the two questionnaires were designed using parallel procedure, the data analysis procedures were the same for both teachers and students. Data from both questionnaires were edited, sorted, and filed on a computer. Frequency distributions were calculated for all the questions items. All percentages were reported as valid percentages with missing data excluded. The obtained data was analysed using SPSS (SPSS stands for Statistical Package for Social Sciences) (Windows version in 19.0). Tables were created in Microsoft® Word.

5.2.4.2 Data Analysis Procedures for Interviews and Focus groups

The raw data, which consisted of the interview and focus groups responses, were processed and transcribed. The data was then organized by grouping all the responses to each question or questionnaire item. The data from each question was then printed out on paper for analysis. The first step was to read and re-read the data in order to develop a detailed knowledge of all responses and to facilitate seeing patterns. The researcher then looked for repetitions and noted these down since this is the basis for devising codes which is a key process in data analysis. Morse and Richards (2002) distinguished between three kinds on coding: descriptive coding, topic coding, and analytic coding. In this study, topic coding was employed. The topics were designated according to the categories previously used in designing the interview and focus groups schedules. The categories were used as preliminary ways of understanding the data. The researcher then looked for patterns across each of the categories, seeking to identify recurrent analytical categories. The transcripts were then grouped and edited again according to the new analytic categories. To analyse data, both inductive and deductive methods were used. Inductive methods were used to explore the data related to the ideas expressed by the respondents, whereas deductive methods were used to analyse the obtained results from the initial data.

5.2.4.3 Data Analysis Procedures for the Observation

Procedures for data analysis based on the observations consisted of the following steps:

- Step One: Global viewing of the lessons.

- Step Two: General coding of all the observed lessons were carried out in the Final Study. Sorting and putting the interaction patterns into the observation schedule for individual teachers.

- Step Three: Detailed description of the organization structures and interactions were developed.

- Step four: Display the relationship between the interaction pattern of the observed lessons and the research questions.

5.3 Results of the Final Study

5.3.1 The Questionnaires

5.3.1.1 The Teacher's Questionnaire (TQ2)

5.3.1.1.1 Demographic Information

Seven characteristics were included in the study. Distribution frequencies based on valid percentages were adopted to compare between the various characteristics. Table 5.5 summarizes the findings:

Items	Variables	Valid percentages
1. Gender	Male	42.9%
	Female	57.1%
2. Age	21-30	42.9%
	31-40	42.9%
	Above 40	14.2%
3. Academic qualifications	´Bac´ certificate	00%
	Licence	100%
	Magister	00%
4. Years of teaching	1-3	42.9%
	4-6	28.8%
	7-9	00%
	Others	28.9%
5. Levels taught	1st year	14.3%
	2nd year	28.6%
	3rd year	57%
6. Teaching sessions per week	1-3	85.7%
	4-7	00%
	Others	14.3%
7. Teaching hours per week	1-3	00%
	4-7	00%
	Others	100%

Table 5.5: The Final Study characteristics of teacher-participants

Table 5.5 shows that 57.1% of the teacher participants were female and 42.9% were male. About 86% of those teachers were less than 40 years old. Only 14.2 of the respondents were more than 40. The sampled teachers were qualified academically. All of them hold a university degree. Among those teachers, only one got a training in special teachers´ Training

Institute. Moreover, the remaining participants started their jobs as secondary school teachers without any specific training. Among the respondents, we could see that some teachers were novice with a short teaching experience, whereas others were experienced. Most of the questioned teachers taught two levels. Those teachers were obliged to teach more than 7 hours per week. This implies that their timetables were overloaded in most of the cases.

5.3.1.1.2 Teachers´ Perceptions of the ATM

In this section, two questions referred to the teachers´ perceptions of the ATM. One question (TQ2-8) referred to the teachers´ perceptions behind the implementation of the new testing policy. The second question (TQ2-9) dealt with the possible teaching methods teachers had adopted as a result of the new model. These two questions used a four-point Likert scale.

1. TQ2-8: Teachers perceptions of the reasons behind implementing the new testing system

TQ2-8 explored the extent to which teachers perceived the reasons behind the implementation of the new testing model. The teachers were given 7 items and had to tick the right answer based on their beliefs, all the questions were designed on a four-point Likert scale, where 1: not a major concern; 2: of little concern; 3: of great concern; 4: main concern. The results are embodied in Table 5.6 below:

What do you see as the major aims of implementing (trying-out) the ATM (ATM)?				
	Not a major concern at all	**Of little concern**	**Of great concern**	**Main concern**
a- To prepare the pupils for their future career.	14.3%	14.3%	28.6%	42.9%

b- **To cope with the present decline in English standards.**	28.6%	42.9%	28.6%	0%
c- **To motivate pupils to use integrated skills.**	14.3%	42.9%	42.9%	0%
d- **To direct pupils' learning.**	0%	14.3%	42.9%	42.9%
e- **To enable pupils to communicate appropriately outside the classroom.**	14.3%	0%	14.3%	71.4%
f- **To meet the instructional objectives set out in the taught syllabus.**	0%	28.6%	28.6%	42.9%
g- **To assess teachers through pupils' performance in the achievement tests drawn out from the ATM.**	0%	42.9%	42.9%	14.3%

Table 5.6: Teachers´ perceptions of the reasons behind implementing the new testing system.

What could be retained from the results embodied in Table 5.6 is that 'to enable students to communicate appropriately outside the classroom' was perceived as the major concern behind the idea to implementing this new testing system. Following this item, three other reasons were seen to be put on an equal foot. Those latter were in order, 'to meet the instrumental objectives set out in the syllabus', 'to prepare the students for their future career", and "to direct the students´ learning'. The two remaining items, 'to cope with the present decline in English', and 'to assess teachers through students´ performance' were perceived by the respondents as not important factors that could lead to introducing the ATM.

Therefore, it could be deduced from the above results that although they appeared to reflect mixed feelings, the teachers´ perceptions of the reasons behind the intervention were consistent with the idea that the teachers believed that this new model could help students in the context under exploration become proficient and competent in English language. Thus, they could be able to communicate using this language with other speakers of it accurately and fluently.

2. TQ2-9: Alterations to teaching methods due to the new testing system

In TQ2-9, there were 14 items, regarding the changes the teachers had made to their teaching possibly in line with the new testing policy. They were listed in Table 5.7, using a four-point Likert scale, where 1: not at all; 2: some extent; 3: a reasonable extent; 4: very large extent.

To what extent have you introduced the following activities in your teaching since the introduction of the ATM?	Not at all	Some extent	A reasonable extent	A very large extent
a- **Employ a new teaching method.**	14.3%	57.1%	14.3%	14.3%
b- **Use a more communicative approach in teaching.**	0%	14.3%	71.4%	14.3%
c- **Put more stress on individual performance.**	14.3%	0%	42.9%	42.9%
d- **Put more stress on co-operative learning.**	0%	0%	28.6%	71.4%
e- **Put more stress on group discussion.**	0%	0%	28.6%%	71.4%
f- **Put more stress on speaking.**	0%	14.3%	42.9%	42.9%
g- **Put more stress on reading.**	0%	14.3%	57.1%	28.6%
h- **Put more stress on writing.**	14.3%	42.9%	14.3%	28.6%
i- **Put more stress on listening.**	14.3%	28.6%	28.6%	28.6%
j- **Put more emphasis on the integration of the four skills.**	0%	42.9%	28.6%	28.6%
k- **Encourage more pupil participation.**	0%	14.3%	14.3%	71.4%
l- **Employ more real-life language tasks.**	0%	0%	28.6%	71.4%
m- **Teach according to the achievement test format.**	14.3%	42.9%	42.9%	0%
n- **Teach according to the "Bac" examination format.**	57.1%	42.9%	0%	0%

Table 5.7: Alterations to teaching methods

From the results included in Table 5.7, it is apparent that items, such as, 'use a more communicative approach in teaching', 'put more stress on co-operative learning', 'put more stress on group discussion', 'encourage more student participation', and 'employ more real-

life language tasks' were high on the list. In addition, the other items like 'employ a new teaching method', 'put more emphasis on the integration of the four skills', and "put more emphasis on listening" were considered to be significant changes that the questioned teachers made to their students because of the new policy. However, the remaining items, particularly the one that referred to 'teach according to the 'Bac' exam format' were not frequently selected by the respondents. This might suggest an alteration in the teachers´ attitudes towards the adoption of the new model, and avoid to be fixed on the 'Bac' exam typology of activities. A practice that was currently carried out with the current testing system. These results clearly indicated that there was a washback effect of the ATM on teachers´ teaching activities.

5.3.1.1.3 Washback on Teaching Materials

There were two questions concerning teaching materials (TQ2-10 and TQ2-11). Items in these two questions were designed on a four-point Likert scale, where 1: seldom or never, 2: sometimes, 3: most classes, and 4: every class. In TQ2-11, 1: not real increase, 2: little increase, 3: some increase, and 4: large increase. The purpose of these two questions was to explore relationships between teaching materials and the implementation of the ATM.

1. TQ2-10: Teaching and learning resources

TQ2-10 was aimed at finding out whether or not the teachers believed that the implementation of the ATM would involve extra teaching and learning resources. These were put in a list displayed in Table 5.8 below:

How do you often use the following resources in your teaching?				
	Seldom or never	Sometimes	Most classes	Every class
a- Textbooks.	0%	0%	57.1%	42.9%

b- Pictures or cards.	0%	42.9%	28.6%	28.6%
c- Data-show (PowerPoint)	0%	57.1%	14.3%	28.6%
d- Magazines.	42.9%	28.6%	28.6%	0%
e- Newspapers.	42.9%	42.9%	14.3%	0%
f- Examination guide.	42.9%	0%	28.6%	28.6%
g- Language laboratory.	85.7%	14.3%	0%	0%
h- Commercial books.	71.4%	28.6%	0%	0%

Table 5.8: Teaching and learning resources

The results included in Table 5.8 indicated that textbooks were the most used resources. Following this item, 'the examination guide' was also chosen by the teachers as a used material. The other items, such as 'picture and cards', 'magazines', and 'newspapers' were thought to be less used. However, the obtained results in this question showed that 'language laboratory' was seldom or rarely adopted by the teachers. Some teachers thought that 'PowerPoint®️ or data show' was used to some extent. Indeed, the latter is the only positive indication that shows that there is a slight washback effect. Apart from this, all the other answers do not imply that the ATM has exerted the expected effects on the teaching and learning resources teachers often used in their classes.

2. TQ2-11: The kind of extra-work practice by teachers

TQ2-11 explored to which extent the incorporation of the ATM had exerted some extra-work in relation to the teaching and learning context. The teachers were given six items, and had to tick the degree of increase of the item from no real increase to a large increase. As to this question, a four-point Likert scale was employed. The results are embodied in Table 5.9 below:

What kind of extra-work or pressure if any do you think the ATM put on you in your teaching?

	No real increase	Little increase	Some increase	Large increase
a- Doing more lesson preparation.	14.3%	14.3%	42.9%	28.6%
b- Preparing more materials for pupils.	0%	14.3%	57.1%	28.6%
c- Revising the existing materials.	0%	42.9%	42.9%	14.3%
d- Adopting new teaching methods.	28.6%	0%	42.9%	28.6%
e- Setting-up new teaching techniques.	14.3%	28.6%	28.6%	28.6%
f- Organizing more examination practice.	0%	42.9%	57.1%	0%

Table 5.9: The kind of extra-work on teaching and learning

It could be seen from Table 5.9, 'preparing more materials for students' increased to a certain extent because of the adoption of the new testing model. Other items, such as 'revising the existing materials', 'doing more-lesson preparation', and 'adopting new teaching Materials'' also increased but with varying degrees. 'To organize more practice on examination' remained a recurrent practice for most of the teachers.

What could be drawn out from these results is that there is a washback effect, but still the questioned teachers found it difficult to get rid of their acquired behaviours. That is, they showed a great resistance to innovation and change.

In summary, the results in the area of washback on the choice of teaching and learning materials (TQ2-11 – TQ2-12) shows clearly that there were no clear effects. Although there was a slight washback effect, but overall practices were not seen to be significantly different from the time that was before the adoption of the new policy.

5.3.1.1.4 Washback on Classroom Teaching Behaviour

Questions related to aspects of daily teaching were investigated in this section. The purpose was to find out whether or not the ATM would influence the teachers´ decisions about, and perceptions of, their teaching activities. In sum, there were three questions: (TQ2-12 -13 -14). One question was designed on a multiple-choice format, and two other ones in a four-point Likert scale, where for TQ2-12, 1: English only; 2: English supplemented with occasional Arabic; 3:half English and half Arabic; and 4: Mainly Arabic. For TQ2-13, 1:

almost never happens; 2: uncommon; 3: common; and 4: almost the whole lesson; and for TQ2-14, 1: seldom or never; 2: sometimes; 3: most classes; and 4: every class.

1. TQ2-12: Medium of Instruction

TQ2-12 explored what the medium of instruction is. As stated previously, the section was designed on a multiple-choice format. The results are included in Table 5.10:

What is the medium of instruction you use when you teach in the classroom?				
	English only	English supplemented with occasional Arabic explanation	Half English and half Arabic.	Mainly Arabic
Valid Percentage	42.9%	57.1%	0%	0%

Table 5.10: Reported Medium of instruction in the context under exploration.

Table 5.10 shows that the majority of the participant teachers used English supplemented with occasional Arabic explanation. One possible reason for this might be teachers´ concerns about their students' general level in English language. Currently, it is believed that students´ general level in English language is not high which allows those students to use the target language at ease. However, 42.9% of the questioned teachers said that they never referred to students´ native language in their lessons. The key aim to proceed in such a way was to avoid any interference between the mother and target languages. Besides, another intention was to make students 'think' in the learnt language, rather than "think" in the native language and communicate with English.

2. TQ2-13: Teacher talk in English lessons

TQ2-13 was related to teacher talk. Teachers were invited to say how often they carried out the following activities: (a) talk to the whole class, (b) talk to groups of students, (c) talk to individual students. Table 5.11 summarizes the results.

How often do you do this in class?				
	Almost never happen	**Uncommon**	**Common**	**Almost the whole lesson**
a- **Talk to the whole class.**	0%	28.6%	14.3%	57.1%
b- **Talk to groups of students.**	0%	28.6%	57.1%	14.3%
c- **Talk to individual students.**	14.3%	14.3%	57.1%	14.3%

Table 5.11: Teacher talk in English lessons

It is noticeable that the great majority of the participant teachers talked to the whole class (57%). 'To talk in groups of students', and 'to talk to individual students' were less common in comparison to item one. As a result, these obtained results suggested that overall teachers still dominated and controlled the classroom talk for most of their lessons. Besides, the results obtained after items 2 and 3 indicated that although teacher talk was dominated by the teachers themselves, the new testing policy seemed to have encouraged teachers to work hard to be more learner-centred.

3. TQ2-14: Teachers' activities in English lessons

TQ2-14 explored how often teachers carried out some activities in their lessons. Five items were suggested. The teachers were asked to choose the appropriate frequency based on their beliefs. The results on this question are reported in Table 5.12 below:

How often do you do the following activities in class?

	Seldom or never	**Sometimes**	**Most classes**	**Every classes**
a- **Explain the meaning of the text.**	28.6%	57.1%	14.3%	0
b- **Explain specific language items such as words or sentences.**	0%	14.3%	71.4%	14.3%
c- **Go through textbook exercises.**	0%	0%	57.1%	42.9%
d- **Organize pair work or group discussions.**	0%	0%	14.3%	85.7%
e- **Organize integrated language skill tasks.**	0%	28.6%	42.9%	28.6%
f- **Talk about achievement test.**	14.3%	57.1%	14.3%	14.3%

Table 5.12: Teachers' activities in English lessons

As to this question, it could be seen that 'to organize pair or group work discussion' was carried most often by the participant teachers. This item was closely followed by three other activities and regarded to have an acceptable frequency of occurrence. These activities are 'to go through textbook activities', 'explain specific language items such as words or sentences', and 'organize integrated language skill tasks' What could be retained from the frequency of occurrence of these activities is that there exists a mixture of beliefs, attitudes, and practices among the questioned teachers. Meanwhile, it sounds that it is very difficult for those teachers to get rid of their traditional teaching methods. Again, what one can draw out as a conclusion from these two different situations is that although there is a slight washback effect of the ATM on the used teaching activities in English lessons, resistance to change is still the prevailing practice.

5.3.1.2 The Student Questionnaire (SQ$_2$)

5.3.1.2.1 Demographic Information

The return rate was 96.6%, who completed and returned back the questionnaires. Five characteristics were included in this section on students' individual characteristics. As what was done in the teacher questionnaire (TQ2), distributed frequencies based on valid percentages were adopted to compare between the various characteristics. Table 5.13 summarizes the findings:

Items	Variables	Valid Percentages
1. Gender	Male	26.4%
	Female	73.6%
2. Level	First year	00%
	Second year	100%
	Third year	00%
3. Branch	Naturel Sciences	00%
	Letters and Foreign Language	100%
	Letters and Philosophy	00%
4. Study hours per week	1-3	7.5%
	4-7	92.5%
	More than 7	00%
5. Preparation hours per week	1-4	97.1%
	4-9	2.9%
	More than 10	00%

Table 5.13: The Final Study characteristics and learning contexts of student-participants.

Table 5.13 shows that 93.6% of the student participants were female and 26.4% were male. All the student participants were selected from secondary classes. All of them were chosen from the letters and foreign languages branch. This choice was based on the assumption that it was easier to control the number of students in this branch because across the Biskra region the number of students who are in this branch is not as big as in the other branches. Besides, these students study English language as a first foreign language. This implies that their devotion to studying this subject is higher than the other students who study English as a second foreign language. 97.1% of the participant students said that the time they spent to revise and prepare English lessons varies between 1 to 4 hours. This result indicates that there is an increase in revision and preparation of the subject for those students, in comparison to the same time they had in the period that preceded the introduction of the new testing system. Consequently, this reality with respect to the aim of motivating students to self-study in English, the ATM has been to some extent effective.

5.3.1.2.2 Students' Attitudes Towards Aspects of Learning and Teaching Abilities Inside and Outside the Classroom

272

Section A: Students´ Attitudes towards Teaching Activities Inside and Outside Classroom

This section explored the students´ attitudes towards teaching carried out inside and outside the classroom by their teachers and by themselves in order to understand possible changes that might have been taking place in the context of the ATM. There were four questions, which were student's questionnaire: Section Two, Part A, questions 6,7,8, and 9. All these questions were designed on a four-point Likert scale of frequency, where 1: usually; 2: sometimes; 3: rarely; and 4: never.

1. SQ2-6: Teacher talk in class

SQ2-6 explored students´ perceptions of how much the teacher talked in their English lessons. The aim was to identify whether or not the introduction of the new testing model reduced the amount of teacher talk and encouraged more communicative approach. The students were asked to grade how often their teachers talk to the whole class, groups of students, and individual students in English lessons. The results are summarized in Table 5.14 below:

How often does your teacher do the following in your English class?				
	Usually	**Sometimes**	**Rarely**	**Never**
a. Talk to the whole pupils.	77.0%	20.1%	1.7%	1.1%
b. Talk to groups of pupils.	40.8%	44.8%	10.3%	4.0%
c. Talk to individual pupils.	30.5%	31.0%	28.7%	9.8%

Table 5.14: Students' perceptions of their teacher talk

From the results embodied in Table 5.14, it could be seen that there is a high 'frequency estimated at 77%, of 'teachers usually talk to the whole students'. Following this item, 'teachers usually talk to a group of students' is estimated at 41.8%. At the end, 'teachers usually talk to individual students' is estimated at 30.5%.

What could be deduced from these results is that the general pattern of teacher talk in class is still in order of, from most to least. This indicates that teachers are still in control and dominance of teaching in class, although the introduced testing system aimed to encourage teachers to be more learner-centred and provide students with more practice opportunities.

2. SQ2-7: Students' perceptions of their teachers teaching activities in their lessons

Students were asked to grade the frequency with their lessons. Students were asked to grade the frequency with their teachers organized in seven different activities. Table 5.15 displays the general pattern of the students´ responses.

How often does your teacher do the following activities in your English class?				
	Usually	**Sometimes**	**Rarely**	**Never**
a. **Tell your class the objectives of each lesson.**	81.6%	12.6%	2.9%	2.9%
b. **Demonstrate how to do particular language activities.**	66.1%	28.2%	5.2%	0.6%
c. **Explain specific language items, such as words or sentence structures.**	81.6%	14.9%	2.9%	0.6%
d. **Explain specific language structures.**	72.4%	18.4%	8.6%	0.6%
e. **Go through textbook exercises.**	69.0%	27.0%	2.9%	1.1%
f. **Assist pupils to answer exercises.**	65.3%	29.9%	8.6%	5.2%
g. **Talk about achievement tests.**	44.8%	20.7%	23.0%	11.5%

Table 5.15: Students´ perceptions of their teachers´ teaching activities

As to this question, we could notice that item one 'the teachers usually tell class the objectives of each lesson' is the most frequent with a percentage that approximates 81.6%. This item is equivalent with item three 'teachers usually explain specific language items'. Following these two items, item 4, 'teachers usually explain specific language structures', is estimated at 72.4%. The remaining activities, such as 'going through textbook', 'demonstrate how to do particular language activities', 'assist students to answer exercises', and 'talk about achievement tests' were less frequent and were seen by the students as activities frequently carried out by their teachers.

It could be seen that there is a slight change of the teachers´ teaching activities in their lessons since there is a variety of activities carried out by teachers. This significantly differed from the time when these teachers were fixed only on grammar and vocabulary. Crucially, this suggests a washback effect of the ATM on classroom activities.

3. SQ2-8: Students' use of English in relation to their classroom activities

SQ2-8 explored the activities the participant students carried out in their classes. There were four activities listed under the question. The results are included in Table 5.16:

How often do you do the following in English class?				
	Usually	**Sometimes**	**Rarely**	**Never**
a. Doing pair work.	68.4%	24.1%	5.2%	2.3%
b. Doing group work.	85.6%	13.8%	0.6%	0.0%
c. Expressing your own ideas in English.	19.0%	60.9%	15.5%	4.6%
d. Asking questions.	21.8%	56.3%	16.7%	5.2%

Table 5.16: Students' use of English in relation to their classroom activities

What is noticeable from the results embodied in Table 5.16 is that activities, such as 'doing group work', and 'doing pair work' are the most frequent with 85.6% and 68.4%, respectively. The two left activities 'asking questions' and 'expressing your ideas in English' were less frequent with 21.8% and 19%.

What could be retained from these results is that because the most frequent activities are 'group and pair works', it is evident that this indicates a change in students' attitudes towards aspects of learning. Consequently, this directly implies that there is a positive washback of the ATM on students´ attitudes to learning.

4. SQ2-9: Students' use of English in relation to their learning activities outside class

275

SQ2-9 explored aspects of students´ learning activities carried out in English outside their lessons. There were five activities listed under this question. The results are reported in Table 5.17 below.

How often do you do the following in English outside class?				
	Usually	Sometimes	Rarely	Never
e. Talk to your teacher in English.	24.1%	24.5%	16.7%	16.7%
f. Talk to your classmates in English.	18.4%	44.8%	24.1%	12.6%
g. Watch TV programmes in English.	57.5%	27.0%	6.9%	8.6%
h. Read newspapers and magazines in English	59.2%	24.1%	7.5%	9.2%
i. Chat on the net in English.	14.9%	25.9%	22.4%	36.8%

Table 5.17: Students' use of English in relation to their learning activities outside class

From the results embodied in Table 5.17, we could see that 'read newspapers and magazines' is the most frequent activity with 59.2%. This item is followed by 'watch TV programmes in English' with 57.5%. The remaining three activities 'talk to your teacher in English', 'talk to your classmate in English', and 'chat on the net in English' were less frequent. According to the obtained results in Table 5.17, difference in frequencies of the occurrence of the five activities give a clear picture of little communication carried out by the students outside the classroom. Ostensibly, this indicates a weak washback effect of the new testing model on the use of English outside class.

Section B: Students' Attitudes towards Aspects of Learning

This section –Student's Questionnaire Section Two, Part B, consisted of two questions with a nominal scale of agreement and likeness. This section aimed to explore possible reasons for the students' attitudes in relation to learning in the context of taking achievement tests drawn out from the proposed ATM.

1. SQ2-10: Students´ motivation to learn English

SQ2-10 explored students' reasons for learning English to see whether or not the new testing policy brought about different perceptions in the students´ motivations. The results are reported in Table 5.18 below:

Do you agree with the following reasons for learning English?		
	Agree	**Disagree**
a. To get a better job	39.7%	60.3%
b. To be able to study in higher education	69.5%	30.5%
c. To be able to communicate with foreigners	90.8%	9.2%
d. To have more and better opportunity in the future	93.7%	6.3%
e. To satisfy my parents' desire	43.1%	56.9%
f. To satisfy school requirements	64.4%	35.6%

Table 5.18: Students´ perceptions of their reasons for learning English

As to this question, what is noticeable from the results included in Table 5.18 is that 'to have more opportunity in the future' was the most motivating reason to learn English (93.7%). In the same vein, 'to be able to communicate with foreigners' was seen to be an important reason to learn this foreign language with 90.8%. An additional high frequency to learn English was that 'to be able to study in higher education', which was estimated to be equivalent at 69.5%. The remaining two reasons 'to satisfy school requirements' and 'to get a better job' were the least frequent.

Consequently, these results showed that the most frequent reasons for learning English are a good indication that the situation of learning that foreign language fits the spirit of the ATM. At this level, it is certain that the new testing policy exerted a positive washback effect on students´ motivation to learn English.

2. SQ2-11: Students´ preferred learning strategies

SQ2-11 explored whether or not the ATM had an influence on students' learning strategies. There were 6 strategies listed under this question. The results are embodied in Table 5.19 below:

Do you like to use the following strategies for learning English?

	Like	Dislike
a.Reading newspapers and magazines in English	40.8%	59.2%
b.Listening to radio programme in English	90.2%	9.8%
c.Taking part in group activities in class in English	91.4%	8.6%
d.Expressing opinions in English in class	90.8%	9.2%
e.Putting more emphasis on understanding English	89.7%	10.3%
f. Doing exercises and homework	89.7%	10.3%

Table 5.19: Students' preferred learning strategies.

From the results included in Table 5.19, it could be seen that the most used strategies by participant students were as follows: 'to take part in group activities in class' (91.4%), 'to express opinions in English in class' (90.8%), 'listening to radio programmes in English' (90.2%), and 'putting more emphasis on understanding English' (89.7%).The remaining learning strategies were less used by the students.

What could be deduced from these results is that the learning strategies that require language production are the most preferred ones. Ostensibly, this leads to say that there is a significant positive washback effect of the ATM on oral communication skills. Certainly, this is what is expected from this new testing policy.

Section C: Students' Attitudes towards Aspects of Testing

This section –student's questionnaire Section Three: Part C- consisted of one question with five statements. All the statements were designed on a nominal scale and used an agreement format. The aim of this section was to explore students´ attitudes towards English tests in relation to teaching and learning.

1. SQ2-12: Students´ attitudes towards tests

Do you agree with the following opinions?	Agree	Disagree
a. Pupils like useful achievement tests	87.4%	12.6%
b. Achievement tests are a valuable learning experience	87.4%	12.6%
c. Achievement tests force pupils to study harder	90.8%	9.2%
d. A pupil's score on achievement tests is a good indication of how s/he will be able to apply what has been learnt	59.2%	40.8%
e. Achievement tests are one of the motivations for the pupils learning	93.1%	6.9%

Table 5.20: Students' attitudes towards tests

Table 5.20 shows that achievement tests based on the ATM stand as a source of motivation for these students to learn. These tests are no longer considered as an operation of grading, but they are seen as a part of the teaching/learning process. In addition, the questioned students regarded this type of tests as a means that 'usually forces them to study harder in lessons'. In one word, these results indicate that 'the new tests are a valuable learning experience'. Contrary to the results obtained in Phase One of the study, the majority of the students in the Final Study said that they 'liked the new achievement tests'.

Consequently, these findings imply that there is a change in students' feelings to, and attitudes towards, achievement tests as they are conceived by the ATM. Students are motivated and willing to sit for them and not anxious or worried to take them as it used to be earlier. On the whole, these attitudes indicate that there is a positive washback effect of the new testing model on students´ attitudes towards aspects of testing.

5.3.2 The Interviews

5.3.2.1 Teacher interviews (02)

This section reports findings from the interview (02) with teachers in the Final Study. When asked about how they have perceived the idea of implementing the ATM, all the interviewed teachers said they have found the idea very exciting. They added; it is very innovative.

T₃ explained that:

I think that the idea of implementing the ATM is innovative for many reasons ; and, I think it is one way to make both teachers and students think more about the importance and the perception of testing.

T₆ added that:

It's a new idea because it is the first time we do something like that; pupils ought to think that it's something for them and it is something sometimes difficult; but, this is at the beginning; later on, they will find it easy.

When asked about the division and sectioning of the achievement tests based on the new testing model, the interviewed teachers emphasized that the format of this model is particularly very useful. They are certain it helps them measure the different skills and language components. **T₂** expressed this belief in these words. He said that:

I found that the format is very innovative. It is holistic, which means one can test everything s/he teaches. The students can be assessed in all the skills because we know that English language is not only reading and writing. Listening and speaking are also very important.

In addition to this short description, **T₈** linked this new format directly to the learners' achievement in English language. She reported that:

The new division and sectioning in this test is a good idea because it shows you clearly the improvement of students' learning. In addition, it encourages them to give more importance to all the learnt skills in this language, and not to be restricted to one or two skills.

T₆ gave her opinion on this format; and, in the mean time, she responded to the view that it is difficult to test listening and speaking. She indicated that:

The new division ... I have found it very well. It is so because we teach our pupils speaking and listening; so, when we test them in listening and speaking, it is easy. It is something not impossible. It is okay. About reading and writing, we usually do that.

T₄ expressed some doubts about the time this new format proposes to carry out the different tasks in the tests. He thought that although the idea of sectioning the test into different sub-tests is good, but still, we have to be very careful about the time allotted for each sub-test. In his words, he reported that:

This division ... I like it, but there is something ... when we use this new model, it takes too much time. 4 hours for one test is a little bit difficult to accept. You know! This time is too long. It can make students stressful every time they are asked to sit for an achievement test. This is a real problem...

Apart from these doubts expressed by **T₄**, the other interviewed teachers welcomed the new division. They did not raise any queries on this new format. For them, it is more effective and practical to adopt.

In order to gain insightful feedback on the teachers' opinions about other issues related to the new testing system, the interviewees were invited to say what they see as major differences between, on one hand, the current testing system, and on the other one, the ATM. To this question, we got different views. For instance, **T₃** thought that:

The ATM is very different from the tests that most students are used to take. First, this model has three sub-tests, which means that the students have the opportunity to be tested on three levels. I mean reading and writing, and listening and speaking ... however, with the traditional testing model, students are used to read only a text and then answer some questions; with the ATM, it is different, they are supposed to listen and build up some skills of English language, which means we cannot imagine someone who is supposed to learn English, just to learn two skills, and ignore the other skills, which are listening and speaking.

T₄ gave a more explicit description on this matter. In his words, he reported that:

There is a huge difference between the two systems. First, the current achievement tests are based on the 'Bac' exam papers: a text, comprehension questions, grammar, and writing. Meanwhile, in the new testing system, there are speaking and listening and other different tasks on other components. I think this is what should be incorporated in a test because a language is not only to read or write, but it is also to communicate, to listen and to speak.

The other interviewed teachers converged to a large extent with the assumptions borne by the two previous teachers. They believed that the new testing policy is more practical and pragmatic since it is not only limited to test a part of language, but transgresses to touch all the skills and aspects a particular language test encompasses. What is also regarded by these teachers to be positive is that the testing of the skills and language components, which were neglected in the traditional testing system, enhances the teaching and learning of these

language aspects, knowing when testing them, both teachers and students would give them more importance in instruction.

Furthermore, on the crucial point of how their students reacted to the implementation of the ATM, the teachers, subject to this interview, provided different responses. For example, **T₃** saw that there are two kinds of reactions. First, those students who were in favour of this innovative act and hence welcomed the change of the traditional testing system. For them, such an idea would allow the students to have a better chance to display how well or badly they are doing in learning English language. On the other side, a second category of students did not see any addition from this intervention since the current testing system does a good job in that it is the best mechanism that prepares secondary students to pass the decisive 'Bac' examination. For those latter, it does not matter whether or not they could communicate accurately and fluently in English language. Rather, what they saw as crucial is to respond correctly to the tasks and get a good mark.

On the same idea, **T₄**'s answer was more straightforward. He said that all his students enjoyed this new model. To prove this claim, he added, with the current testing system, he usually felt that his students did not feel at ease and seemed to be anxious the days that precede the test. This feeling had completely disappeared with the incorporation of the new achievement tests. The students showed enthusiasm to undertake the tests. He expressed this atmosphere as if the students were going to 'attend a party'. They were not worried of the test, and did not show any signs of disgust. Instead, they were happy and comfortable. **T₂** shared the same assumptions with **T₄**. In precise words, he reported that:

The students reacted positively to the new testing system model because they found they could be tested in all the skills of language: listening, speaking ... and this could give them an idea about their level in learning English.

Contrary to the previous views of the interviewees, **T₈** thought differently. She advanced that that situation was pretty complicated, especially, at the initial stage of the implementation of this new testing policy. Students, she added, showed a tough resistance to change. They did not accept that their teachers employed a new testing mechanism to assess them, particularly when they were told that the tried out new achievement test was not officially recommended. Nevertheless, this negative attitude turned to become positive when these students got familiar with the model, and felt that it bore interesting and attractive things that they were not accustomed to. This experience was also shared with **T₆**, who indicated that her students became more motivated and willing to sit for the new achievement tests with time. She also said that she is certain that this positive attitude can take a wide dimension if the model is officially adopted by the educational authorities and becomes the current testing system in the Algerian secondary schools. In this case, teachers and students would be compelled to use it and the benefits are guaranteed.

In sum, what could be retained from the interviewees' responses to this question is that the majority of the teachers, who tried out this new testing model, welcomed this innovation and appreciated to get a new test with a different format and varied contents. They agreed to say that such a way of assessment enables them to obtain very insightful feedback on instruction. Although this echoed sentiment is too early to make a definite conclusion on the new testing model policy, but in itself it is a good indication that these teachers reacted positively to the new intervention.

5.3.2.2 Students Interviews (02) and Focus Groups (02)

This part of the research will present findings from both students' interviews and focus groups in the Final Study:

When asked about how they perceived the ATM, the majority of the interviewed students found this new testing system a good idea and an innovative experience. Although their first impressions showed that they were still discovering the secrets of this testing mechanism, but their answers demonstrated that they appreciated and welcomed it. To arrive at this conclusion, the students made a comparison between the current testing system and the newly introduced one. A general agreement among those interviewees indicated that there exists a huge difference between the two. For most of them, they trusted the evaluation of the new tests, no matter what type of judgement they got at the end bad or good. Besides, one of the most striking things observed in the interviewees' responses in the interviews and focus groups was that the feelings of boredom and anxiety completely disappeared, and were replaced by feelings such as likeness and joy to undertake tests. Some significant expressions to this first question in the individual interviews and focus groups are collected and displayed below:

- *actually, I felt the ATM is a very good testing system to show what I know and I don't know!* **S₁**

- *it's a good idea to tell the pupils their true level.* **S₆**

- *it's a good way to see the level of students; it helps us to improve our capacities in English language!* **S₈**

- *it's a good test and I find it really interesting because I can now precisely know my abilities in all the skills of this foreign language.* **S₉**

- *I like this model because I found its questions not so complicated and also not so hard to answer. Actually, I appreciate it. It tells me my level in English.* **S₁₁**

- *For me the ATM is a new idea. It's the first time I see it.* **S₁₆**

- *I see the ATM is a good idea because the contents of the test correspond to what I study in class. I like it.* **S₂₂**

On the question to express their opinions about the format of the ATM, the students' answers were almost similar to say that the new division and sectioning of the achievement

tests based on the new model were comprehensive and consistent. In the students' words, these views were expressed in this way:

I think the new division of the ATM is very interesting. It's great; it's this kind of tests that gives me more chance to answer the questions. It's better that the tests we used to do. S₁₂

The format of the ATM is very clear. It's totally different from the tests we usually had. The new one is made-up of three sub-tests, whereas the traditional one is only one test. The difference is here because testing is not writing and reading; it's all aspects of language. S₁₃

S₂₄ converged with all the interviewed students, his answer is a conclusion to these students' responses. In precise terms, he said that:

The new testing system is a good idea. I see that it's a test which can develop the students' skills and abilities. It's a test which enhances learning. It's different from the other traditional tests. In addition to reading and writing, it assesses listening and speaking. I think in this way, testing can influence positively both teaching and learning.

Concerning how they perceived the distribution of tasks per sections, items per sections, the allotted time for each sub-test, and the scoring scale, the interviewed students showed some misunderstandings of these test-aspects. The students' answers were limited and not thorough. They indicated that they needed time to get familiar with the new tests to have sound judgements on these issues. Interestingly, this does not imply that the interviewed students had negative attitudes towards these aspects related to the new testing model; rather, the discussed questions might have been pitched at too high level for them to grasp at their age. Moreover, it is certain that the reason of this problem could be attributed to the students' general linguistic background, which largely hampered them to express their ideas well in English language, knowing that the individual interviews and group discussions in the Final Study were carried out in English language, and not in Arabic. That is why, on this question, the interviewed students answered only using short phrases or simply telegraphic and fragmented sentences, such as 'yes', 'no', or 'great', and 'interesting'.

To the other crucial question, 'how do they perceive the marks they obtained after sitting for the ATM?' There seemed to be a mixture of attitudes towards this matter among

the interviewees. For a great deal of these students, the new model made them feel satisfied about their marks no matter what these marks were, high or low. They thought in such a way because they saw that their marks are true and not misleading as they used to be before the intervention. However, for some other interviewees, and certainly they were not so many, they believed that their marks after undertaking the new achievement tests are not real. To justify this belief, the interviewed students went back to the high pressure the new tests exert on them, which in many times make them underperformed in the tests. For these students, such a pressure is mainly due to the three sub-tests that make-up one achievement test, and also to the time they often take to answer the questions.

At the end of the interviews and the focus groups with students in the Final Study, the participants were asked to express their views on whether or not they are in favour to adopt this newly testing system, or simply reject it. The responses of the interviewed students indicated that the majority of them welcomed this new policy, and advocated that it would be better to keep this new testing model. Surprisingly, some of the respondents suggested that the ATM is not an end in itself, but it is more important to keep revising and adjusting the contents of the test to make it more effective. In order to attain this goal, the interviewees' answers provided some evidence that this new policy should go further to be encouraged by the educational authorities, and adopted by all the secondary school teachers in the local scene. Their motive to this claim is mediated by the idea that given the importance placed on the new testing model, the intensity of motivation to learn English increased among them. Students think that they become more interested in the benefits of the language and the desire to communicate with English language inside and outside the classroom. What mattered more for them is to transgress the idea of exam-preparation classroom to concern more powerful motivation to learn language for communication and to achieve success in their studies.

5.3.3 Classroom Observation Outcomes

The following section in the study reports the findings of 4 from the 8 teachers observed in the Final Study.

5.3.3.1 Participants Organization

Table 5.21 summarized the interaction patterns of the lessons with T_1, T_2, T_3, and T_4.

Number of Teacher T1	Percentage of lesson time				Total
	Whole class T to S/C	Whole class	Group	Individual	
	65%	00%	25%	10%	100%
T2	65%	00%	25%	10%	100%
T3	25%	05%	50%	20%	100%
T4	50%	05%	35%	10%	100%

Table 5.21: Participants organization on lesson for the Final Study

The findings from the observed classes in Phase Three in this study indicate two crucial remarks. First, 3 from the 4 teachers observed (T_1, T_2, and T_4) spent the majority of class time talking to the whole class since the calculated time for these 3 teachers is estimated to be equal or more than the total time of class work. Individual work is still low, although there is a slight increase devoted to that category of work in comparison to that of the Preliminary Study. The rationale behind spending a considerable amount of class time talking to the whole class, according to detailed observation and based on interviews with the teachers, is that it was difficult for the teachers to get rid of their teaching practices where they dominated class discussion. In the meantime, the same group of teachers showed a great resistance towards change and innovation from a total dominance to activities where most of the work is carried out in groups or individuals. However, it is worth mentioning that among these teachers and within this teaching atmosphere, T_3 seemed to be an exception since this observed teacher succeeded to reduce the time spent for whole class discussion to devote much of her work to group activities. Unlike the other 3 teachers, she spent 25% of her class

time only to whole discussion. 50% of the time is spent to group work, and 20% to individual tasks. It seems that the purpose of this change towards group and individual works was to provide students with practice opportunities. For her, it is believed that group work helped students learn from each other and gave them more practice opportunities, especially the oral.

Moreover, the second remark under this category demonstrated clearly that in comparison to the results obtained about the same issue from Phase One (see Table 3.46) there is a considerable change in the behaviours and practices of the observed teachers. For instance, in Phase One, apart from T_3 whose calculated time devoted to group work was estimated to be 40%, this time had interestingly increased for almost all the other teachers in this last phase of the study. Respectively, the percentages shifted from 0% to 25% for T_1, and from 4% to 35% for T_4. It should be noted that this category was treated differently to what was carried out before the ATM implementation wherein students were involved in discussions or in conducting group tasks. The purpose was to help them interact and communicate efficiently using English.

5.3.3.2 Activity Types and Content as Percentage of Class

The analysis of the observed lessons with the 4 teachers presented in Table 5.22 showed (a) what types of activities were carried out in the lessons, and (b) who was holding the floor and in what ways.

288

Activity type	T1	T2	T3	T4
Teacher activities:				
Lecturing	20%	20%	05%	20%
Explaining	25%	15%	10%	20%
Directing	20%	15%	00%	10%
Student activities:				
Listening	00%	05%	05%	05%
Reading	05%	05%	00%	05%
Speaking	05%	05%	05%	05%
Writing	00%	05%	00%	05%
Grammar	15%	15%	05%	05%
Vocabulary	05%	05%	05%	05%
Group work	05%	00%	50%	10%
Pair work	00%	05%	10%	05%
Oral work	05%	05%	05%	05%
Total	**100%**	**100%**	**100%**	**100%**

Table 5.22: Classroom activities of all the four observed teachers as a percentage of class time from phase three

It could be seen that T_1 was still the predominant focus of the class. 65% of the lesson time was spent in teacher activities; the remaining 35% was spent on reading texts, explaining grammar, and some oral tasks. For T_2 and T_4, 50% of the lesson time was dominated by the teacher activities: lecturing, explaining, and directing. The findings in T_2's and T_4's showed that there is a slight interest in student activities. With T_2 even though grammar activities took 15% of lesson time, listening, pair work, and oral practice shared another 15%. For T_4, the same activities were estimated to be 20%, bearing in mind that this increase in these activities is compared to the results observed in Phase One.

Contrary to T_1, T_2, and T_4, T_3 significantly focused on the student activities. In precise terms, only 15% of the lesson time was devoted to teacher activities. This directly implies that this teacher reduced the time that the other teachers made in total or partial control of teaching. Instead, T_3 spent 85% of lesson time to the student activities. The big amount of this time is addressed to group work. The estimated time on this particular student activity is estimated to be 50%. Obviously, this result indicates that the teacher's strategy has turned from teacher-centred to the learner-centred approach. In itself, this is a good indication of communicative teaching. Similar to this finding, T_3 spent 10% of her time to oral work. This

means that through this activity this teacher encouraged her students to participate and interact using English language. The remaining time is shared by the other activities, such as listening, speaking, and pair work.

To summarize the findings from Table 5.22, a cross-comparison of the percentage of the time on students' activities in the lessons observed after the implementation of the ATM showed that there was an increase in the time devoted to student activities. Certainly, the results differed from one teacher to another; nonetheless, it seems that the majority of the teachers involved in this study worked to make their lessons more interactive and the adopted approach more learner-centred. For these teachers, this new strategy seeks to provide more opportunities to practise English language inside and outside the classroom.

5.3.3.3 Teaching Materials

This section summarizes the findings related to the use of teaching materials in classroom teaching. It was found that the most used material by the 4 observed teachers was the official textbook. This result is similar to some extent to what was seen on the same issue of teaching materials from Phase One. However, it is worth mentioning that in addition to this general finding, other materials were observed to have been employed by the teachers. For instance, **T₃** used PowerPoint and slides while she carried out-group work tasks. The students were usually invited to watch an extract from a documentary, often chosen with a great care from multimedia sources like internet, and were asked to fulfil some relevant tasks. In the case of these situations, the teacher did the job of a guide and did not interfere or intervene so much, giving the students freedom to perform what they were asked to do. Thus, it was noticed that the students, reducing the amount of the lesson time given to the teacher, did most of the work. Moreover, the use of multimedia was also part of the listening and speaking activities. Being currently exposed to authentic audio-visual materials aided the students to train their ears to native language; and this in itself yielded these students an

opportunity to try to speak fluently. Other teachers were observed to have used some teaching materials, such as cards, pictures, magazines, and newspapers. These particular materials were adopted mainly in lessons teaching reading and writing.

In what follows, a discussion of the findings from the four data collection methods in the final stage of this study will be presented in the next section.

5.4 Discussion and Summary of the Findings

The findings of this study from Phase Three obtained using four data collection methods have led to a better understanding of the participants' reactions at the micro-level of the washback effects of the ATM on teaching, learning, and assessment practices. First of all, a review of the questionnaire findings from both teachers and students provided insightful feedback regarding classroom practices. With regard to the teacher's questionnaire exploring the teachers' perceptions of the reasons behind the intervention, it can be seen that those latter were consistent with the underlying theories on which the new testing policy rested. Teachers are aware that the main purposes of the introduction of new tests are in line with the idea that considers the students in the context under exploration as active agents who possess a high degree of competence of the learnt language. This, in itself, is the fundamental purpose of teaching English language in the Algerian secondary school, which is clearly identified in the three syllabuses at this level, and that enables students to communicate to some extent fluently and accurately, or at least, with some ease, inside and outside the classroom. Concerning the alterations to teaching due to the intervention, evidence from the teachers indicated that these participants no longer are fixed on a given typology of activities that are currently included in the 'Bac' exam papers, as it used to be in the period of time that preceded the introduction of the new testing system; rather, there is a more positive attitude of these teachers towards the relative importance of classroom activities, and the subsequent use of new teaching methods. In this respect, it was found that the participant teachers in this

291

study give a higher weighing to the activities of 'putting more stress on the co-operative learning', 'putting more stress on group discussion', and 'putting more stress on speaking and the integration of the four skill'. Certainly, these activities and perhaps other ones can increase students' communicative competence, indicating a degree of interest of these teachers to make changes in certain aspects of teaching. In short, there seemed to be a washback effect of the new policy on aspects related to teachers' reactions to the new system, the reasons behind introducing the ATM, and alterations to teaching methods due to the new policy.

Contrary to the positive indications obtained with the teacher' perceptions of, and reactions towards the new policy, the result in the area related to aspects of teaching materials revealed that the teachers still rely on the official textbook as the major resource in their teaching. It appeared that the majority of the teachers relied mainly on the content and organization of general English textbook as the source of their plans because they do not want to deviate from the official instructions recommended by the educational authorities, mainly those of the general inspectors, bearing in mind that one of the limitation of this study was that it is extremely difficult to secure the necessary support and involvement of all the teachers in this study because it is often perceived not to have the authority and status to compel the participants to apply all what it suggests. Given this reality, it is not surprising that the washback effect on classroom teaching resources appears to have been minimal since the teachers for the most part did not try to change so much the teaching materials, although in many cases they expressed in their answers they were not convinced of what they were doing, and they showed some interest to opt for other resources.

In terms of classroom teaching behaviours, the majority of the teachers used half English and half Arabic as the medium of instruction, according to the teachers this was due to the weak general level of their students in English language.

292

Furthermore, although it was noticed in this study that teachers became aware to talk less to students, and talk more often to groups or to individuals aiming at providing the students more practice opportunities, talking to the whole class was still the main activity in the classroom. The results of the teacher questionnaire in the study showed that the pattern of teacher's talk to the whole class was the most frequent, though, as discussed earlier; the new policy did appear to encourage teachers to be more learner-centred. In brief, there seemed to be a washback effect of the new testing system in the area of teacher, but it was not as considerable as it was expected.

Regarding the factors that influenced how teachers perceived the issue of assessment and evaluation, the results indicated that these factors proved to be significantly different from the time when the teachers' main expectations were to obtain good scores in the tests, and pass the 'Bac' examination to a new time when the teachers' expectations are to obtain satisfaction in teaching. Accordingly, there is an evidence that using the ATM as an agent of change could possibly make teachers pay more attention to their own expectations, as well as to their students.

When aspects of learning were investigated from the teachers' perspectives, it appeared that these teachers were more willing to encourage their students cope with the innovative act, rather than change their own methodology. Teachers' attitudes towards how students should study in the content of the new policy was positive to put more emphasis on aspects which suggest change; the same attitude was obtained towards learning strategies. It appears that teachers' attitudes towards aspects of learning in terms of learning aims and learning strategies indicated that there is a washback effect, even though the teachers' reluctance to make change could be explained by the resistance to innovation and the short time the new policy has taken to be discovered and well understood by the teachers in the context under investigation.

Finally, in the teacher questionnaire from Phase Three, exploring teachers' reactions to the new testing model shows that generally there is a positive washback. A great deal of the participants welcomed the intervention. They were attracted by the new format and contents of the achievement tests based on the ATM. They declared that they are ready to abandon what they used to use as assessment instruments, and embrace the new testing system. Such a positive reaction towards the policy is also interpreted in terms of the fact the new model encompasses a new philosophy, approach, and methodology.

Besides washback on teaching, major issues regarding washback on learning explored in this study were the basic concerns of the student's questionnaire in this last phase. First, with respect to washback in relation to the student' attitudes towards teaching and learning activities inside and outside class, the results obtained of this aspect confirmed the findings from the teacher's questionnaire. The general atmosphere is characterized by a dominance of teacher control. Although one major aim of the introduction of the ATM was to encourage speaking and oral practice, and to enhance pair and group work, there was a lower frequency in carrying out these activities and practices. Again, this reality should not hide the fact that in comparison to what prevailed in the scene during the pre-implementation period of the new testing system, the situation has to some extent changed but not to what is expected. Perhaps, there is a need for more time to get positive feedback. Overall, the result indicated that there was actually a minimal washback effect on the activities that students carried out inside and outside class.

With respect to the students' attitudes towards aspects of learning, the analysis of the results showed that the students' expectations correlate with the new intentions of the new testing system implementation. The students' response claimed that to have better opportunities in the future and to be able to communicate with foreigners in English language, this language could be a ticket to the world beyond the context where they live and

study. The students also revealed an acute consciousness of the fact that English is a language that they are compelled to learn to open new horizons and get a place in the world of today. This general mood corresponds to a great extent to the new policy in which the functional use of English language is encouraged to be considered as an important aspect of learning. In addition, in terms of the students' preferred learning strategies, it is certain that with regard to these students perceptions of the new model, their answers reported a positive reaction, and influence to the new achievement tests since the most used strategies were the one that suggest a washback effect on oral communicative strategies.

With regard to students' attitudes towards the impact of the new policy on aspects of testing, the findings indicated a positive shift in the students' perceptions. It can be seen that before the intervention, all the students said they did not enjoy sitting for the test. This feeling of disgust was interpreted in how these students viewed tests. For most of them, a test used to be equivalent to a sanction. After the incorporation of the new model, the students turned to regard the new achievement tests as a useful means to see how well or badly they have achieved in correspondence to what they learn in class. More importantly, this new perception considered tests as effective instruments to learn English language. In sum, these students viewed the new model in a positive light. In particular, this indicates that they have changed their attitudes towards the function of tests.

The analysis of the findings related to aspects of the ATM format and contents indicated that the students appreciated the new structure, considering that sectioning a test into these sub-tests allows a better assessment of the four skills and language components. Besides, the students believed that the variety and richness of the contents of the tests could be a good indication that they are representative to a large extent to their lessons. This new attitude towards this matter completely differs from that of the pre-implementation period. Finally, the findings in this section confirmed this positive attitude towards the new testing

model since it seems that all the students welcomed the new scoring procedures and scales. They thought that those latter are comprehensive since they can provide them with actual final outcomes. All in all, it is ostensible that the changes that occurred to the format and contents of the new, proposed tests have brought about positive attitudes of the students towards both assessment and evaluation.

Secondly, besides to the findings and discussion of the results from the teacher's and student's questionnaires, a discussion of the classroom observation outcomes is indispensable to understand the nature of washback in the context under investigation. Because the major purpose of the new testing policy was to improve students' communicative skills, it was essential to explore a sample lesson in detail to see whether there are any discernible washback. In terms of participants' organization, the patterns of teacher interaction in the classroom revealed some differences among the 4 observed teachers. For 3 of them, teacher talk dominated the lesson time, whereas with 1 teacher, there were some differences, especially, in allocation of time in teacher and student talk. Indeed, in T_3's lessons, there was less teacher talk than in the other teachers' lessons. T_3's lessons were much more learner-centred, and consequently showed the most amount of student activity. As a result, the introduction of the ATM seems that it has encouraged a more learner - controlled approach, even though this result was observed with only one teacher. This in itself is a positive light.

With regard to the activity types and content as percentage of the class, the classroom observations revealed that although in general the approaches to teaching did not vary, since it was observed that the teachers resisted to keep on using the same practices they were carrying out for a long period of time, there was an increase in the time devoted to student activities with serious attempts to make the lessons more interactive. It is apparent that because of the new policy, the teachers have become aware that it is crucial to change many of their classroom activities, and the content of their teaching to achieve more substantial

296

results at the end. What is more significant on these aspects of activity types and teaching content is that there were many signs of a change in the teachers' fundamental beliefs and attitudes about how to carry out teaching, and what should be included in lessons to make all the participants in the teaching operation get profit. In brief, what was notable from the study, even though it was a short period of time since the new policy had been in operation and short period of research, there existed an actual washback effect of the ATM, or a perceived effect, indicating the power of the existence of the new tests.

In terms of the final aspect in the observations, which mainly concerned the used teaching materials, it is actual that almost the same conclusion drawn out from the previous aspects of the activity types and contents of lessons was attained. The findings suggest that the observed teachers most of the time relied on the official textbook, but with a progress in the application of the new policy, the teachers began incorporating some other new teaching resources that pertain to the kind of teaching materials that correspond to where less teacher talk and more oral student practices were involved. Examples of these materials included CD's, data show, pictures, cards, and other audio-visuals. Overall, it can be seen that a washback effect of the ATM has motivated the teachers to review at least some of their teaching materials, and suggest new ones.

Thirdly, a further careful look at the interviewed teachers' and students' responses revealed how these participants perceived the new policy. All the interviewees agreed to say that the tests are completely different from the current. Essentially, they have found that the new crucial criterion these tests have brought is that they do not eschew any skills or language components; reasonably, they thought that because in one test, three sub-tests that assess the four skills, vocabulary and grammar make it possible to provide students with actual grades. In particular, a number of the teachers affirmed that this new model has made it clear for them what stages one has to follow to be systematic, and what to consider when time

297

comes to develop their tests. This finding aided the teachers to avoid relying on the same tests that frequently looked similar to the 'Bac' exam papers. Finally, the findings from the interviews revealed that the intended washback, as it is expected by this study, was to an extent achieved, despite the fact that the new policy is still at its initial phase of diffusion and implementation.

Conclusion

This chapter has discussed and summarized the findings in Phase Three, which comprised the investigations conducted after the introduction of the new testing model at the micro-level in the context under exploration. To carry out this exploration, four data collection methods were employed. Statistical analysis were utilized to examine the findings from the questionnaires. Indicative and deductive methods were used to analyse the data related to the ideas expressed by the participants in the individual interviews and focus groups. The COLT scheme based on Part A was employed to study the information collected through classroom observations. A review of the discussion of the findings in this chapter revealed that a washback effect of the new model occurred regarding some teaching and learning aspects, whereas at the level of some other aspects, washback was superficial or even absent. It is essential to point out that the short period of time of the new model had been in operation, and other external factors hindering the intervention could explain the obtained results. What did result, however, the findings of this phase have led to a better understanding of what participants think of the ATM, and its effects on teaching, learning, and assessment practices.

Chapter Six: Conclusions of the Thesis, Implications for Pedagogy, a Review of the Washback Model in this Study, and Limitations

Introduction

This chapter culminates the research in the present study. First, section one synthesizes the findings which answer the research questions raised in the introduction of this thesis. Following the present synthesis, implications for key participants and other parties, directly or indirectly involved in the study, are suggested. The implications are meant to provide some pedagogical recommendations that can stand as important insights that would be useful for test reform projects in the future. This section will draw a set of conclusions across the three phases of the research by re-visiting the major issues in this washback study. In the third section, the exploratory washback model, examining this phenomenon designed to fit the Algerian educational context, will be reviewed. Last, the chapter ends by discussing some limitations of the study.

6.1 Synthesis of the Findings in the Present Study

The present section revisits the major issues and tries to investigate the findings from the three phases of the study. The aim of this exploration was to investigate the impact of a particular new testing model in a specific educational setting, and to see whether it was possible to extract from this experience insightful feedback which could help us to understand that a plan to innovate through testing is realistic in the Algerian educational context. The major topics in this study consisted of an examination of the context where the research took part before and after the incorporation of the new project. The research paid much more attention to the washback effect of the suggested testing model on both teachers' and students' perceptions of, and reactions towards, this new system. It is evident that the scope of the study was limited to the micro-level because of the nature and purposes of this

study with the intention to understand how the main participants within the Algerian educational context identified areas of washback intensity in teaching and learning and reacted to changes made in their current testing system.

The investigation required a critical review of washback as it is presented in general education and in language assessment, and a survey of a number of important ideas in the literature of educational innovation. The review of the literature on this central research concept indicated that washback could be either positive or negative. Based on the findings in the literature review on this phenomenon in this study, an attempt was made by developing a framework to analyse a language teaching innovation. One on the fundamental factors that is greatly considered in this framework was to describe the situation on the ground and how it looked like before the introduction of the innovation. Thus, it was necessary to gain a good understanding of the situation before attempting to introduce any educational change.

Before the introduction of the new testing system, the first remark that can be gleaned from the discussion of the results of this study is that the current testing system did exercise an influence on what was going in the classes. This impact was found to have been primarily mediated through the 'Bac' exam guide. Given this situation, it was a current practice that EFL teachers in the Algerian secondary schools regarded the 'Bac' exam manual to be the source on which teaching and assessment rested. This particular model prescribed and guided the teachers and became the course syllabus by default as a determinant of the content of teaching and classroom assessment leading to undesirable washback effects on teaching, learning, and testing practices. But why did the teachers use the 'Bac' exam guide the way they did? Why did they deviate from the official methodological recommendations of the official syllabus? The answers to these questions could be found in the results obtained from the teachers' responses and behaviours in the Preliminary Study in this research.

One reason that the teachers put trust in the 'Bac' exam guide in terms of content could be because they saw concrete manifestations of the decisive examination in the secondary school cycle present in the contents of the used manual; and therefore, they found it easier and more pragmatic to opt for a teaching strategy that mostly relied on this exam-book. In addition, these teachers had seen that most of the texts and tasks that are included in the guide might appear in the 'Bac' exam papers. Thus, this might have led these teachers to follow faithfully the manual because its content usually resembles to what the final examination encompasses.

Equally so, the teachers had felt that this was what students expected and this would help them prepare for the decisive exam successfully. It is also noticeable that the policy of the schools required from the teachers to obtain high scores in the 'Bac' exam to judge them to be skilful and competent. Due to this requirement, the teachers spoke about the pressure they were under to get their students succeed, and the feelings that the 'Bac' exam limited the work they should do with their students. Those teachers used the term "narrowing the curriculum" to describe this phenomenon. For others, their major aim turned to become 'passing the 'Bac' exam', and not to teach English for some precise and definite instructional objectives well identified in the relevant syllabus.

The analysis of the students' results produced a similar picture, it can be seen that the students also experienced negative washback effects of the 'Bac exam on the strategies they used to learn English language. Under some special circumstances, the students came to see learning this foreign language in instrumental terms. They become more interested in the utilitarian benefits of language rather than to become proficient and competent in it. Their immediate goal of learning this language was to achieve success in the 'Bac' exam. What mattered more for these students was to obtain a good score in this exam and be able to follow their studies at the university level rather than learning the language efficiently. In this

sense, it was obvious that the instrumental motivation was more powerful for learning English language than the purpose to learn it to be able to use it appropriately inside and outside the classroom. It is likely that once the examination is over, there is very little to sustain this kind of motivation.

In addition, negative washback effects of the current testing system on the students was found to create an interplay between this phenomenon, motivation, and anxiety. It seems that because of the dominance of the exam-oriented methodology, the general impression gave place to an abandoned feeling of dissatisfaction created by the intense focus on, and preparation for the exam. The students did not enjoy their classes. They felt tired, uninterested, and experienced boredom. This led to a lack of attention, participation, and discipline during the lessons. Under these circumstances, one could argue that boredom could turn out to be a factor likely to hampering learning. There is some evidence that the poor performance of the students in the achievement tests made them feel demoralized and demotivated. As a consequence, self-esteem plummeted, constructing themselves in terms of the level to which their performance was ascribed. The findings are interesting because they indicated that a negative washback was present.

In summary, both the teachers and students in the period of time preceding the incorporation of the new testing model generally viewed the current testing system in a negative light. They also shared the ideas that the achievement tests used in the Algerian secondary schools were useless. To reiterate, these participants recognized these tests were conducted without being based on any theoretical background or following systematic procedures. More importantly, they did not fit for their intended purposes. The result was that they provided unreal scores, and hence made inaccurate influences. With regard to this conclusion, it is worth noting that the obtained findings using the different data collection

methods in this period of exploration answered the raised questions and confirmed the proposed hypothesis.

Having synthesized the results of the study at its initial stage, we shall review now what the present study has provided us as evidence about exam washback after the implementation of the ATM.

In the period of time that followed the introduction of the new testing system, the study explored whether the implementation of the new tests had any washback effects on classroom teaching in terms of teacher and students behaviours. The investigation was carried out in relation to the participants' reactions in the research content, and further, examined the effect of the change on aspects of teachers' and students' attitudes and reactions to the new testing policy. The methodology used in this part of the research aimed to capture the reality, variation, and complexity of changes in classroom practice of the participants.

It is notable that the analysis of the results after the incorporation of the new achievement tests yielded positive insights because of the washback effect of the new testing model on teaching and learning. It was clear that most participants were pleased that the new tests were different from their predecessors and they matched the desired instructional objectives. Thus, the teachers saw that this innovation prevented them from limiting their teaching to only certain features of the curriculum. They advocated that the content and format of the developed tests based on the new model persuaded them to adopt a more comprehensive teaching methodology. Likewise, the students, felt that the content of the new tests were relevant to assess their abilities that make them use the appropriate foreign language in everyday life. They thought that to write their answers, they needed to make use of their knowledge and mental capacities instead of just recognizing the correct answers from a set of options, often using guessing. They also declared that things became clear for them

because the instructions and the scoring scales were no longer confusing as they used to be in the case of achievement tests on the traditional testing system.

In addition, evidence from both teachers' and students' findings showed that although the ATM was not compulsory in the Algerian secondary schools, the teachers turned to place a great deal of importance on the skills and activities that were neglected in the period of time that preceded the introduction of the new system. There was a recognition that the two important skills listening and speaking needed to be taught and assessed. One of the main reasons of this was so that it would force students to work on the development of these skills.

The new model had some impact on the content of teaching, during both the ordinary teaching period, and the examination preparation period. The teachers spent time to teaching listening and speaking. During this time, these teachers switched from relying on using limited part of the textbook to cover the textbook activities and exercises for these two skills. Another positive influence concerned the impact of the new testing instrument on how teachers design and develop classroom tests. The teachers claimed that this new model urged them to become more principled and regard the process of assessment as a systematic operation rather than a mere practice by which they usually came to assign scores to their students.

In short, from a review of the findings, the new testing system had positive features which made it an improvement over the previous current testing system. It was powerful enough to encourage teachers to pay equal attention to the four skills, use student-oriented techniques, and make use of innovative ideas in their teaching.

Unlike these positive claims, the study also revealed that the proposed new testing model had little impact on some areas particularly that of the teaching methodology used. Indeed, some of the teachers practices indicated clearly these teachers continued to employ the teacher-centred approach that they had traditionally adopted rather than the learner-

centred activity based teaching recommended by the suggested language teaching approach in the Algerian school system. A possible explanation for this was that these teachers found it very complicated to get rid of their teaching practices because of the short time since the adoption of the model. It would have been also difficult for the teachers in this situation to introduce new teaching techniques that play down the used methods in favour of language skills which encouraged them to let students work things out for themselves. Teachers would have needed a lot of support, over time, to be able to understand the ideas underlying the principles of innovation and also to make their students ready for challenge.

These were further factors at the school level which made it difficult for teachers to respond positively to the new policy. These teachers were not committed to change since the new ideas about teaching and assessment were suggested by someone who did not have the authority to impose it. For some other teachers, it was also difficult for them to let go of the traditional role of teacher as transmitter of knowledge, and to understand that they were supposed to be helping their students to develop skills rather than aiding them in attempts to learn specific information. This might have been the result of a lack of confidence in their abilities, and this would have made it more intricate for them to pass confidence on their students.

It can be seen that the factors were inhibitive ones. They made it difficult for some teachers to put the actions brought by this innovation into practice. The result was that these teachers generally spoke that they were in favour of the new testing policy; however, the attitudes they professed to have did not always translate into the kind of behaviour that was intended and did not necessarily indicate a change in their basic beliefs about assessment. The teachers accepted the content of the new tests but they did not adopt the recommended methodology. This would count as superficial change and the impact was instrumental rather than conceptual.

305

The above situation suggests that the attitudes towards teaching and learning that participants brought with to the classroom might have a greater impact on the how learning experience is constructed. Evidence from the participants' findings suggest that these participants' approach to teaching and learning were influenced more by beliefs brought to the lessons than by the demands of the test. In some explicit terms, course content was clearly affected by the introduction of the new testing system, but any influence on teaching and learning methods was less obvious and was mediated by participants' beliefs.

To emphasize, the ATM had an impact on a great deal of participants in this study. It increased their awareness in the roles they can take in contributing to positive changes in assessment practices. It provided them with the opportunities to critically re-examine their past and current practices in assessment. It made them discover the problems of the weaknesses they were currently doing. It let them learn theoretically and practically knowledge that offered them directions in improving these problems and weaknesses. Crucially, they became aware of how they can make use of this knowledge in the future. Another important gain that emerged from this study for the participants is that the new testing model strengthened their confidence in making sound judgments on their students' achievement and progress. This would make the teachers more self-consistent in assessing the students' performance. In short, this testing system has succeeded to a great extent to change the participants' attitudes, and behaviours in teaching, learning, and assessment practices.

To summarize, the findings from the period of time that followed the implementation of the new model confirmed the proposed hypotheses in this phase of the study.

6.2 Pedagogical Implications

The pedagogical implications take the form of several suggestions. These are addressed to several key participants, and others that are identified in this study, or might be concerned in other researches similar to the present one in further studies in the hope that they could lead to improving the chance of washback to work in a beneficial way. Nevertheless, it should be made clear that the suggestions are not meant to be exhaustive or definitive. Rather, the practicality and effectiveness of each suggestion needs to be explored later on.

6.2.1 A Useful Achievement Test Implications

It is crucial for EFL teachers to bear in mind that the idea of the existing 'ideal' or 'best' language test is not actual. In many cases and contexts, such a belief has led to the emergence of a number of misconceptions which constitute a strong affective barrier to many practitioners in the field that often end in providing unsound judgments and inferences on the final outcomes. Because of this reality, there is a need to clear-up the terrain by identifying these misconceptions and their resulting problems. And, more importantly, propose right assumptions on what should be to avoid these misconceptions. For this reason, in our study in this part of the research, several suggestions can be put forth for the teachers enabling them to design and develop useful achievement tests preparing their students for being evaluated and assessed for the intended purposes in the right way. Thus, some of these suggestions are summarized below, and include the following ideas:

- Be aware that there is only one test for any given situation.

It has been proved that the idea that there is a set of procedures a teacher can follow to develop a test that would be the 'best' one for some given purposes and situations is a wrong belief. It has thus been argued that believing in such a way often leads to inaccurate inferences.

307

- Avoid misunderstanding the nature of language testing and the purpose of language test development.

It is assumed that if teachers do not pay attention to the idea that the nature of language testing is different from one category of tests to another, this will result in tests which do not meet the specific needs of the purposes of test takers. And thus, these tests will be inappropriate for their targets.

- Realize that there is a need to have reasonable exceptions about what language tests can do and what they should be.

It had been common that if teachers´ unreasonable exceptions about what language tests can do, and what they should be, this certainly will result in the fact that teachers will become frustrated when one is unable to find or develop the appropriate test.

- Avoid being intuitive when developing a language test.

It is obvious that when a teacher bases himself/herself only on his/her intuition to develop a language test, this often leads in realizing unsystematic procedures that often yield monotonous tests. More dangerous, these developed tests can stand as a bad source on which one can draw relevant inferences which can mislead anyone who wants to make sound judgments.

- Be careful in using the language tests that are usually proposed in commercial books.

It is argued that the use of tests that are from commercial books can make teachers lose the ability to rely on themselves to make a test that corresponds to their teaching. Very often, these tests are typical examples of the large-scale exams that are meant to train the candidates how to respond to given questions. In many cases, these do not serve the assessment of students in the course they study.

- Avoid placing blind trust and faith in typical models of tests.

It has been shown that when teachers place a blind trust and faith in some models, for instance, the "Bac" exam in the case of our study, this leads to losing confidence in one's own capacity for developing and using tests, as well as feeling that language testing is something that only "experts" can understand and do. It is believed that avoiding some of the misconceptions described above will help teachers to be competent in language testing and also be able to develop a set of reasonable expectations for the provision of useful tests. In addition, this will give them a sense of confidence in their own knowledge and skills in this endeavour. Thus, competence in language testing involves the following:

a. An understanding of the fundamental considerations that must be addressed at the start of any language testing efforts;

b. an understanding of the fundamental issues and concerns in the appropriate use of language tests;

c. an understanding of the fundamental issues, approaches, and methods used in measurement and evaluation;

d. the ability to design, develop, evaluate, and use language tests in ways that are appropriate for a given purpose, context, and test takers; and

e. the ability to critically read published research in language testing and information about language tests in order to make informed decisions.

It is obvious that teachers are not compelled to be experts in the field of language testing. Rather, being competent in this area simply implies that these teachers should possess some knowledge and develop some skills that can help them to develop language tests on the ground of correct basis. This itself will allow the attainment of valid inferences and make the process of evaluation more valuable. Thus, the approach adopted in this research does not

limit itself to the use of some technical terms in language testing; rather, it transgresses to concern two fundamental principles which will provide important assumptions:

a. a theoretical grounded basis for developing and using language tests; and

b. understanding that tests will enable teachers to make their own judgments and decisions about either selecting or developing a language test situation.

6.2.2 Implications for Teachers

Several considerations can be put forth for EFL teachers in the Algerian secondary schools. Those latter concern aspects related to the form, format, and content of achievement tests. Once again, the basic aim of these considerations and suggestions is to make this type of tests fit for the intended purposes, which are particularly to evaluate the achievement and progress of students in accordance to the taught syllabuses and their relevant instructional objectives. In this sense, what to consider is that:

- The teacher has to identify as clearly as possible the purpose of the test. Generally, this concerns knowing how students are progressing, finding their weaknesses and strengths, working out ways helping them, and using information from the assessment for future planning of classes.

- Through the above activity, the teacher should be able to explicit the specific uses for which the test is intended. It involves clearly eliciting the inferences s/he intends to make on the ground of test results, and any specific decisions s/he will make should be based upon these inferences.

- The teacher has to rely on clear theoretical bases when designing his/her tests. The advantage of basing a test on clear theoretical background will definitely allow the teacher to become competent in developing and using achievement tests for the intended purposes.

- The teacher should be able to design and develop an achievement test in a way which is appropriate for a given purpose, content, and particular students. For sure, basing a test on clear theoretical principles allows the avoidance of random, intuitive, and ready-made tests that are often irrelevant to instruction and which provide inconsistent and inappropriate inferences about the final outcomes.

- The teacher has also to design a test that encourages and enables students to perform at their highest level of ability.

- The teacher should be creative and flexible, and avoid being monotonous and fixed on the same test each time s/he is asked to develop one.

- The teacher should develop as clearly as possible the language abilities and skills to be taken into account when designing his/her test. By defining the language abilities/skills in this way, the teacher is making clear the definition of the constructs to be measured.

- Another consideration is to decide which specific components of language abilities in the syllabus are to be included in the construct definition. In many cases, the teacher wants to make inferences about some specific components of some language abilities and may thus define the constructs in terms of those components. This might be the case if the test is to measure the degree of achievement of specific syllabus objectives. The teacher will then most likely base the construct definition on the specific components of language abilities that are to be included in the syllabus.

- The teacher has to consider the different learning styles of his/her students. Such a consideration is often determined by the importance allowed to the cognitive/psychological abilities and makes-up of students. This can be

311

made possible on the ground through the design and development of achievement tests that fit the requirements of learning objectives, particularly those that are stated in the common Bloom's taxonomy.

- The teacher has to consider the different learning styles of his/her students. This allows the elimination of all sources of frustration and anxiety. A positive attitude vis-à-vis the test will help achieve better scores and hence enable the attainment of substantial progress in learning English language.

- The teacher should be careful in considering some social and cultural contexts when developing his/her tests. For instance, privacy and confidentiality, which are viewed by many as fundamental rights of students, should be carefully considered when designing and developing tests.

- The teacher has also to be usually aware of the importance of considering some critical test qualities and characteristics, such as: validity, reliability, practicality, interactiveness, authenticity, and impact. The teacher needs to strive to achieve an equilibrium among these particular qualities.

- The teacher should be selective in choosing the sample of data to be tested on both quality and quantity; an equilibrium must be reached between what has been learnt/taught and what will be tested.

- The teacher has to make the kind of feedback that the students receive relevant, complete, and meaningful to the content of the syllabus. This allows the confirmation of the students' own perceptions of language ability.

- The teacher should not neglect the students' characteristics when designing and developing achievement tests. This mainly concerns personality

characteristics, background knowledge, aptitude for learning, native language, and some other matters. The involvement of these characteristics will allow students to perform at their highest level in that target language.

- The teacher should make available all the required human and material resources. S/he has also to seek for more assistance and co-operation with other teachers. An assistance of some sort is usually needed to perfect the final test.

- The teacher has to treat students as responsible individuals by providing them with as complete information as possible about how a test is developed and scoring procedures followed.

- The teacher must consider the testing process as an integral part of the teaching process. S/he should emphasize on the importance of professional judgment and commitment to enhance learning. The relationship between teaching and testing should be regarded as that of partnership.

- The teacher has to know in detail the requirements of each stage in the test development process. More importantly, s/he should also consider the decisions and repeat the tasks that have been performed at another stage.

- The teacher has to be aware of organizing the different procedures in the development stage that help to monitor the usefulness of the test.

- **What to suggest?**

In addition, a number of suggestions are also put forth for the teachers. These include assumptions, such as:

- The teacher should be careful in choosing the appropriate texts that have to be congruent with the taught programme.

- The texts should be of medium length; generally, the number of words in one text should vary from 170 to 200 words. The length should be in complete correspondence with the allotted time for the tests.

- Students should be familiar with the style of the language of these texts. It should not be beyond the general language abilities of the students.

- It is appropriate that the teacher adds a list of glossary, which explains some of the words included in the text. Preferably, this list is added at the end of the text. This step enables students access to the general idea of the text.

- The teacher has to mention the source of the text, whether the text is of his/her own production, an extract from another one, or simply taken from the textbook.

- The teacher should specify the vocabulary in the text. It is more appropriate to use a text that comprises an amount of lexical-items from the ones that were dealt with in the class.

- Since the aim of achievement tests is to assess and measure the degree of mastery of vocabulary in relation to the taught programme, it is recommended test lexis independently from the text. This allows the testing of the required vocabulary store.

- Before developing the test, it is important to specify the structure taught in class. The teacher has to choose which one of these structures has to be included in the test.

- What is crucial on this last point, it is essential to vary the structures from one test to another in order to avoid the fixation only on the same structure.

- It is of a great value to vary the test methods from one test to another. This permits to develop the cognitive abilities of students and enhance their learning through testing.

- It is important that teachers have to be aware that the purposes of achievement tests are well defined. They markedly differ from those of the "Bac" exam. In this sense, they should not be totally fixed on the "Bac" exam guide when time comes to develop their tests. Rather, they have to work to be creative, innovative, and flexible. This will minimize the negative washback effect of this standardized examination.

- The teacher has also to be aware that the scoring procedure is an integral part of test development. This is why s/he should be explicit in adopting clear and well-principled methods.

- Besides, s/he has to inform her/his students about the scoring procedure. This will create a positive atmosphere between the teacher and the students and vice-versa.

- Finally, the teacher should not use the same scoring method. S/he has to vary it according to the criteria s/he sets out at the beginning of the development of tests.

6.2.3 Implications for Students

Several assumptions could be stood as recommendations for the students in this study. For instance,

- Findings from this study should prove a helpful source of information for students about how long they might realistically take to make specific gains from achievement tests.

- Findings also will need to pass beyond the immediate requirements of tests if they are to be adequately prepared for academic study in the learnt language.

- Students should not obey to the permanent pressure that external examinations, particularly the "Bac" exam, exert on them and that often leads to negative washback effects.

- Students need to know and understand the limitations of this external examination, and seek opportunities to go beyond the demands of this examination.

- Students should also be aware of the existing relationship between learning and testing. The evidence in this study points to the strong tie between these two processes.

- Students should ask their teachers for clarifications if they are having difficulties and problems with what they study in the classroom.

- Students should feel free to raise abjections whenever the over-use and emphasis is on the commercial exam materials, or if they see that their teachers use other materials that do not meet their expectations.

- Students should express feelings and dislikes more openly to teachers. Such feelings include feelings towards tests, the kind of tasks a test comprises, unclarity of the scoring procedures, and other similar matters.

- One of the most important findings in this study is the relationship between emotional regulation and students' performance. Therefore, understanding the processes that are included in a test-taking situation is very indispensable for students.

- In addition, this study has shown that tests by their nature are stressful encounters. Accordingly, it has been concluded that taking into consideration the affective side when developing an achievement test would enhance test performance of students.

- The value of this study on this aspect has provided specific information on how to guide students when it is time to develop an achievement test.

- Students' performance may be enhanced by the reducing of anxiety through making the developed achievement tests reflect to a large extent what the students study in the target language in class.

- The implementation of the new testing model in this study would be beneficial to student if the lessons could be extended to cover the teaching of the specified abilities and skills included in the contents of the syllabus. This means that focus on only parts of these syllabuses would not help students to develop all these skills and abilities.

- It is recognized that generally students need to be provided with a variety of additional learning resources and these students should be explicitly informed about them. In this sense, learning resources other than test-oriented ones are a valuable contribution to students' learning.

- Students should be given with additional opportunities and assistance. This makes these students become more strategic and independent learners.

- Students need to carry activities that are based on multimedia. It has been found that these students are likely to be more interested in learning English language if they learn in pleasant surrounding using the activities and strategies they like.

- To engineer positive washback, it is important to understand students' degree of achievement and progress and set tasks at an appropriate level, so that the tasks may be responsibly challenging to motivate students perform at their best.

6.2.4 Implications for Educators

Although not involved directly in this study, since the main focus of the investigation was carried out at the micro-level, including both teachers and students, there are some implications for educators who have a link with teaching and learning. These include mainly people such as, general inspectors, teacher trainers, test developers for large-scale tests and examinations, and perhaps some other ones.

- Educators should explain the nature of language tests in general, and achievement tests in particular during the ordinary in-training services that are frequently held in the school years, or in seminars and workshops often organized for teachers to train them in some testing methodologies.

- Educators need to go beyond the superficial knowledge that most EFL teachers have on the subject of testing, and make these teachers get in-depth knowledge and assumptions on language tests, the two processes of test design and test development, and how to make these tests useful.

- Educators should not forget to enlighten the teachers with some test phenomena that are current in the assessment practices. One of this would be washback. Along with the identification of these phenomena, educators need to help the teachers know how to achieve the positive side and implications of such practices.

- In order to allow teachers develop useful language tests, educators are asked to tell the teachers how to proceed to realize clear and detailed

specifications. Training these teachers on this matter makes them self-relied to develop their own tests, and avoid being dependent on external examinations, as we recommended in the part devoted to teachers.

- Educators should make teachers aware of the importance to revise and update their tests. This makes them not very fixed on the same typology of tests.

- Educators need to urge teachers to conduct washback studies. These studies should not only cover the impact of the test design and procedures, but also the attitudes, perceptions, and practices of teachers. In the same vein, particular attention should be paid to students who are the main beneficiaries of the tests.

- The study results should be disseminated to key parties within the education system so that decisions can be made about teaching, learning, and assessment practices.

6.2.5 Implications for Researchers

Based on the experience accumulated from this washback study, several practical recommendations could be made to researchers who are interested in conducting research on this subject. For example:

- It is important to review the washback studies that were carried out in different contexts. From these studies, the researcher needs to draft examinations which can be scrutinized and presented in a form of a full checklist of features which have been researched, and areas which need further exploration. This examination specification makes it easy for evaluators to know where to start a new study in this field.

- It is valuable to theoretically ground the washback study. This means that the researcher should adopt a framework when carrying out investigation into the feasibility of test projects. It is essential to analyse the obtained information on the basis of the used framework in order to determine whether change is desirable and how much change should be attempted.

- It is advisable to adapt the Henrichsen's framework since it has been proved throughout the available literature review on this matter that it is the most comprehensive and practical framework. The researcher can proceed to some modifications in this framework to make it suitable to the context and the nature of the investigation.

- It is appropriate that researchers should look very critically at the experience of previous reformers before deciding on the feasibility of a project. This procedure enables them to get a transparent picture on the subject matter before the introduction of the new reforms.

- It is crucial to go beyond the participants at the micro-level (teachers and students) and consult representatives, such as: policy makers and examination designers at a macro-level to make the expected test reform more effective and substantial. It has been noted that limiting the scope of the investigation in the participants at the micro-level cannot help the change spread and convince other parties that it is what should be done.

- In order to make the desired impact more considerable, the researchers need to establish research links with other researchers on the same domain, organizations or research groups, and centres of research. These contacts can help the researcher get insightful feedback on what s/he is researching and hence make his/her subject more understandable.

- It is advisable in further washback studies to avoid carrying out this type of studies individually; instead, the researcher should propose a management team in reform project. This team might include in addition to teachers, general inspectors, teacher trainers, curriculum designers, and perhaps other people. This team could ensure that the project is academically respectable.

- It is essential to pilot and try-out the proposed testing project before its implementation. This is especially important when the components of the new system encompass different structures or contents. This makes possible to test out the flexibility of the new ideas and procedures to gather feedback from all the participants in this project.

- Constant evaluation of the new testing system should be carried out from the earliest stages of the introduction of the new tests until the end. The obtained results should be analysed by the researcher to decide whether the design and procedures were appropriate or not. Some remedies, adjustments, and modifications can be made to the new testing system to improve it.

- It is more appropriate not to be too idealistic or ambitious when introducing new testing systems into the educational system. It is expected that too much resistance to any change or innovation can occur. An important issue to consider is that the researcher should be patient before seeing his/her objectives become true.

6.3 A Review of the Washback Model in this Thesis

The results of this study suggests that there is sufficient evidence to lay claim to the applicability of the washback model used in the present exploration as a framework for proving whether or not a washback effect has occurred on teaching and learning. This model,

which was described in Chapter 1, Section 1.6, represents the mechanism by which the effects of washback from the new designed achievement tests impact on classroom teaching. In this respect, washback is seen to involve participants, processes, and products within the Algerian educational context.

Because the nature and scope of the washback model in this study was limited to capture the picture of the washback effects on the participants' perceptions of, and attitudes towards, the introduction of the new testing model, there is evidence to say that the obtained results confirmed the underlying assumptions in the suggested hypotheses. That is, "a useful achievement test for EFL secondary classes in Algeria will influence positively teaching and learning; and conversely, an achievement test for EFL learners that is not useful will influence negatively teaching and learning".

In the line with the same conclusion, the other hypotheses were also confirmed, but this time with less washback effect intensity since the hypothesis, which proposes that "a useful achievement test will influence attitudes towards the content and method of teaching and learning", proved to be right with the content of teaching and learning, and very simplified with the teachers' methodology. Above all, although there were some areas in the washback model that need to be reviewed to make them more effective, this model in this study related successfully test design issues to participant attitudes, beliefs, and behaviours.

It is apparent that from evidence provided above that further refinement of the washback model is required, particularly in the domain of washback on the final outcomes of participants. A subject which is intricately related to the implementation of the new testing system. Future research in the local scene and in the same context as it was in this investigation might actually seek explanations from participants for the relationship between their beliefs and attitudes to the test and final outcomes.

To emphasize, washback is a complex phenomenon. This investigation has shown that to study this phenomenon, there is a need to tackle many context variables relevant to the educational system in which the test is used. Also, this exploration has confirmed that it is obligatory to rely on a suggested washback model. It depends on that this model could be adopted from the available literature review, adjusted from the existing models, or simply an invention of the researcher. The advantage of using a washback model is to draw a picture of the washback mechanisms with respect to the effectiveness of implementing new tests and how these latter influence the participants' beliefs and behaviours in accordance to teaching, learning, and assessment practices.

Drawing out from the above assumptions, it is obvious that not only the new testing model in the present study has potential for changing the content of teaching, but also it has limitations. In other words, while this study has suggested that a change in test design is necessary condition for alterations in the teaching and learning practices, it has been indicated that this is not fully made possible. Tests cannot influence teachers to change their practices if they are not totally committed to the new ideas and also if they do not have the skills that will enable them to experiment with, evaluate, and make appropriate adjustments to new teaching methods.

6.4 Limitations of the Study

The first limitation of this study pertains to the nature of the research. The present study is a washback study that evolved over time, so a longitudinal research would have been an appropriate strategy to capture and monitor the ebb and flow of the introduced model's impact. Ideally, the researcher would have liked to observe the participants for a long period of time that could go beyond the time associated with a doctoral thesis. However, this was not possible in a view of the available time. The data for this study were collected in two successive school years, so it is preferable that further researches on this project will take

much more time to meticulously investigate the raised questions and obtain in-depth insights from the proposed model.

The second limitation is the presentation of the data in the thesis. Although the researcher collected a wealth of information to sustain the claims made in the thesis using four data collection methods, space and time limitations inherent in a doctoral project did not allow the researcher to present all that was collected as data, and go into as much depth in presenting them, as might have been possible and preferable. For instance, the data from the questionnaires results, particularly those that were analysed through SPSS, were limited to minimum and were not all displayed. However, this rich and unused data could be a useful resource for future publications.

The third limitation is about the researcher's status. It was extremely difficult to secure the support and involvement of all the participants in a doctoral study because it is often perceived not to have the authority and status that is conducted by official educational deciders. For this reason, the researcher is particularly grateful to the teachers, students, secondary school headmasters, and some general inspectors, who devote effort and valuable time to participating in this study. It is believed that the findings would have been more interesting if the study had more teachers in Biskra region which had to try out this new testing model.

The final limitation of this study relates to the scope of the research. It is transparent that the researcher has claimed that this washback study in the Algerian educational context was only limited to participants at the micro-level (i.e., teachers and students), and did not concern other participants at the macro-level, such as: decision makers and Examinations Authority in Education Department in the Ministry of Education. As this is perhaps the first empirical washback study in Algeria, the researcher decided to pay more attention to key participants at the micro-level that would constitute a solid ground to understand how things

are going on at this level. The other reason of limiting the scope of the study would be because the researcher found it very complicated to carry out an individual research at a very broad scope that could go beyond teachers and students. It would have needed more energy and resources to realize it.

Conclusion

To conclude, this final chapter in the present thesis linked the findings of the study to the research questions raised throughout the different phases in this investigation. Firstly, the research questions were revisited in turn, summarizing the evidence from the various stands of the study. Following this discussion and synthesis, implications for participants suggesting pedagogical recommendations were presented. In addition, an evaluation and a review of the washback model in this study were displayed. These confirmed the notion that washback is a complex phenomenon, and that tests have power to lead for a change to take part in the teaching and learning; but, in the meantime, they have also limitations. Thus, it is too much to expect them to succeed if there is absence of an authority to impose the introduction of a new testing model. Above all, it is hoped that the ideas discussed in this chapter will provide general guidance for researchers who wish to introduce a change through tests, and that the implications for test development will prove particularly useful.

General conclusion

To reiterate, what urged us to choose this research subject is that when observing secondary school teachers in Algeria, we find that their assessment practices do not provide them with specific information about their students' actual performance. It is believed that one of the major reasons of this problem is that the testing approaches, methods, and instruments available in the field are not appropriate to achieve comprehensive judgement and make accurate inferences. In practical terms, the developed tests are currently ready-made language tests, or copies of tests from the 'Bac' exam papers, or merely tests developed by the teachers in a hurry or by intuition. In more explicit words, these tests do not rely on any well-defined theoretical background. They do not assess the developed skills and competencies in instruction. They do not cover a considerable amount of the contents of the used syllabuses. They are not developed through systematic procedures. More dangerous, these tests are not related to specific and general educational goals. In some areas, they can be useful tools for evaluating parts of the programme, or certifying general competencies, but less helpful in diagnosing students' needs and strengths in the different stages of the learning process.

In this respect, one of the fundamental purposes of this study is to sensitize the secondary school teachers in Algeria that the deleterious effects of the current testing system can be overcome by switching to the evidence that there should be permanent revision and revising of the ways that these teachers usually employ to check their students' achievement and progress in English language. They need to realize that testing is not a simple operation of assigning scores, but it is an on-going process that requires test design and test development stages and that the final outcomes involve building a logical case in support of a particular interpretation. Because of this way of conceiving the issue under study, in this investigation we proposed an ATM that would be regarded as the essential basis for quality

control throughout the entire test development process. We would argue that in this study, all test development and use should be informed by this model. The most important consideration in it is the use for which it is intended, so that the most important quality to develop achievement tests based on this model is its usefulness. In this way, the proposed ATM is thought to make the teachers in the context under exploration aware of the possibility to have their tests closely linked to the goals of instruction as to be almost indistinguishable from them; it is asserted that teaching to these tests, as they are conceived by the model, would be considered a virtue.

Methodologically, because this investigation examined the high complex phenomenon of washback of the ATM on teaching and learning, it used an empirical study. The latter drew on theories and research undertaken in the field of language education in general and language assessment in particular. The exploration was more about the nature of exam washback particularly from teachers' and students' perspectives, thus, addressing the call for a mixed-methods approach advocated by research methodologists and language testing specialists to be the most appropriate in the case of this study. To attain the intended purposes of the research, the latter employed both qualitative and quantitative methods of enquiry throughout as evidence and has gone into great detail in designing, piloting and validating the instruments of analysis used to calls made in the literature.

Significantly, the longitudinal nature of the present study added a dimension in how to collect data in somehow a long period of time. This markedly differed to the most of other washback studies that were carried out in other contexts and that were limited only to collecting data using some instruments that did not consider the evolution of the expected change. It is believed that this way of tackling the subject is an efficient research strategy since it makes one able to follow the events of change and hence become also able to capture a transparent picture on the studied phenomenon. In precise terms, this investigation

highlighted the role of teachers as potential actors that can make testing more important, not, as stated before, a simple operation of scoring students. In the same sphere, the ideas discussed in this thesis showed that students are equally important in washback process, not as beneficiaries of tests, but also because their attitudes and feelings towards the test are important ingredients in the learning process. In addition, this study revealed that there might be a concern of the consideration of other participants in the teaching operation, particularly those who are usually seen to form the macro-level of teaching and learning process, because of their role in the presence or absence of washback. Certainly, this gives a clear insight of the big picture.

Therefore, drawing on from the discussion of the results obtained after the implementation of the new testing model, we have come to consolidate the idea that testing is not an easy matter. Indeed, the present investigation has strongly confirmed that testing is a systematic procedure of gathering data about general and specific abilities from performance on tasks designed to provide a basis for consistent and comprehensive expectations of learners' outcomes. In this study, we have conceived assessment as a process through which one needs to conceptualize, develop, and put into practice a final useful test. We have realized the dire need to consider the nature of available research findings on testing and rely on classroom experience to guide test design and development. We have assumed test-content as an experience of test development and looked at how this content may be determinant through the process of test specifications. We have also asserted some conventional response formats as the test option to the category of tests we intend to design and develop. We have emphasized on the importance to make serious efforts to develop test-tasks and test-items that assess specific dimensions of the students' achievement and progress. We have understood more and more than one must be aware that test procedure should measure the students' behaviour and cognitive processes that have been proved to be

desired outcomes in teaching. We have advocated that teachers must conceive testing as a congruent part of teaching and learning. And finally, we have claimed that tests must be considered as a source of collecting useful feedback as how to improve instruction.

All in all, and referring back to all what has been dealt with in this thesis, the ATM should not be regarded as an end in itself, but it has to be looked at as a starting point to remedy the array of anomalies pinpointed in the current testing system and that frequently result in a multitude of irreparable inadequacies that negatively impact the teaching and learning processes. It is believed that one way to become competent in assessment practices is to better understand fundamental issues and approaches in measurement and evaluation. At the end, if we are certain that the proposed testing model is neither the ideal nor the best instrument; we think that it could be taken as an attempt that helps teachers in Algeria to make their tests useful. Thus, this modest contribution can urge teachers to approach testing without prejudices and put it into positive use. The challenge to those involved in this area will be to ensure that the situation can be improved with more research and findings.

Hopefully, this study can serve as a guide for future washback studies, particularly in the local scene, where it has been observed that researches of this type are very rare. The research remains open to other studies in the Algerian educational context in the following years.

References

- Adelman, C., Kemmis, S., & Jenkins, D. (1980). Rethinking Case Study. In H. Simons (ed.), Toward a science of singular. (pp. 45-61). Centre for Applied Research in Education, UK: University of Eastern Anglia.

- Alder, P. A., & Alder, P. (1994). Observational Techniques. In N. K. Denzin & Y.S. Lincoln (eds.), Handbook of qualitative research, 337-391.

- Alderson, J. C., & Wall. D. (1993). Does Washback Exist? Applied Linguistics, 14, 15-29.

- ………., ..., ..., & Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A study of washback. Language Testing, 13 (3), 280-297.

- Allen, P., Fröhlic, M., & Spada, N. (1984). The Communicative Orientation of Language Teaching: An observation scheme. In J. Hanscombe, R. A. Oren, & B. P. Taylor (eds.), On TESOL' 83: The question of control, (pp.231-252). Washington, DC. TESOL.

- Allwright, D. (1983). Classrom-centred Research on Language Teaching and Learning: A brief historical overview. TESOL Quarterly, 17, 191-204.

- …………, … (1988). Observations in the Classroom. London: Longman.

- …………, ..., & Bailey, K. (1991). Focus on the Language Classroom. Cambridge: Cambridge University Press.

- Andrews, S. (1994). The Washback Effect of Examinations: Its impact upon curriculum innovation in English language teaching. Curriculum Forum, 4 (1), 44-58.

- ………., …. (2004). Washback and Curriculum Innovation. In L. Cheng, Y. Watanabe, & A. Curtis (eds.), Washback in language testing: Research contexts and methods, (pp. 37-52). Mahwah, New Jersey: Lawrence Erlbaum Associates.

- ………., …., Fullilove, J., & Wong, Y. (2002). Targeting Washback: A case study. System, 30 (2), 207-223

- Bachman, L. F. (1997). <u>Fundamental Considerations in Language Testing</u>. Oxford: Oxford University Press.

- ……….., .., …., & Palmer, A. S. (1996). <u>Language Testing in Practice</u>. Oxford: Oxford University Press.

- ……….., .., …., & ……….., .., …. (2000). <u>Modern Language Testing at the Turn of the Century</u>. Accusing what count counts. In J. C. Alderson, & B. editors, language testing 17 (1).

- Bailey, K. M. (1985). <u>Classroom-centred Research on Language Teaching and learning</u>. In M. Celce-Maria (ed.), Beyond basis: Issues and research. TESOL Rontledge: Newbury House.

- ……., ..., ... (1989). <u>Washback in Language Testing</u>. Princeton, New Jersey: Education Testing Service.

- ……., ..., ... (1996). <u>Working for Washback: A review of the washback concept in language testing.</u> Language Testing, 13 (3), 257-259.

- ……., ..., ... (1999). <u>Washback in Language Testing.</u> New Jersey: Educational Testing Service.

- Baker, E. L. (1991). <u>Alternative Assessment and National Policy</u>. Paper Presented at the National Research Symposium on Limited English Proficient Students' Issues: Focus on evaluation and measurement, Washington, DC.

- Biggs, J. B. (1993). <u>The Quality of Teaching and Learning in Expanding Tertiary System</u>. New Horizons, 34, 102-104.

- Brindley, G. (2008). <u>Educational Reform and Language Testing</u>. In E. Shohamy, & N. H. Hornberger (eds.), Encyclopaedia of language and education, 7, language testing and assessment, (pp. 365-378). New York: Springer.

- Brown, J. D. (2000). <u>University Entrance Examinations: Strategies for creating positive washback on English language teaching in Japan</u>: Testing and evaluating. SIG Newsletters, 3 (2), 4-8.

- Bryman, A. (1989). <u>Research Methods and Organization Studies</u>. London: Routledge.

- Buck, G. (1988). <u>Testing Listening Comprehension in Japanese University Entrance Examination</u>. JALT Journal, 10, 12-42.

- Caine, N. A. (2005). <u>EFL Examination Washback in Japan: Investigating the effects of oral assessment on teaching and learning</u>. Unpublished Master dissertation, the University of Manchester, England.

- Campbell, D. T., & Stanley, J. C. (1963). <u>Experimental and Quasi-experimental Designs for Research on Teaching</u>. In handbook of research on teaching. Chicago: Rand McNally.

- Campbell, D. T., & Fiske, D. W. (1969). <u>Convergent and Discrimination Validation by the Multitrait- Multimethod Matrix</u>. Psychological Bulletin 56, 81-105.

- Cavan, S. (1977). Review of J. D. Dougla's (1976) <u>Investigative Social Review: Individual and team field research</u>. The American Journal of Sociology, 83 (3), 809-811.

- Chaudron, C. (1988). <u>Second Language Classroom: Research on teaching and learning</u>. Cambridge: Cambridge University Press.

- Cheng, L. (1997). <u>How Does Washback Influence Teaching? Implications for Hong Kong</u>. Language and Education, 11 (1), 38-54.

- ……., ... (2004). <u>The Washback Effect of a Public Examination Change on Teachers' Perceptions towards their Classroom Teaching</u>. In L. Cheng, Y. Watanabe, & A Curtis (eds.), Washback in language testing; Research Contexts and methods, (pp.147-170). Mahwah, New Jersey; Lawrence Erlbaum Associates.

- ……., ... (2005). <u>Changing Language Teaching through Language Testing: A washback study</u>. Cambridge: Cambridge University Press.

- ……., ... (2008). <u>Washback, Impact, and Consequences</u>. In E. Shohamy & N. H. Hornberger (eds.), Encyclopaedia of language and education, 7, language testing and assessment, (pp. 349-364). New York: Springer.

- ……., ..., & Watanabe, Y. (2004). <u>Context and Methods in Washback Research: The influence of language testing on teaching and learning.</u> New Jersey: Lawrence Erlbaum.

- Chinda, B. (2009). <u>Professional Development in Language Testing and Assessment: A case study of supporting change in assessment practice in in-service EFL teachers in Thailand</u>. Unpublished Doctoral thesis, Nottingham University, England.

- Cohen, A. D (1994). <u>Assessing Language Ability</u>. Boston: Herrile and Herrile.

- Cohen, L., Lawrance, M., & Morrison, K. (2005). <u>Research Methods in Education</u> (5th edition). Oxford: Oxford University Press.

- Cooper, R. (1989). <u>Language Planning and Social Change</u>. Cambridge: Cambridge University Press.

- Creswell, J. W. (2009). <u>Research Design: Qualitative, quantitative, and mixed-methods approaches</u>. Thousand Oaks, California: Sage Publications.

- Curtis, A. (2000). <u>A Problem-Solving Approach to the Management of Change in Language Education</u>. Korea TESOL journal, 3 (1), 1-12.

- Davies, A. (1985). <u>Follow Any Leader: Is that what language tests do</u>? In Y. P. Lee, A. C. Y. Y. Forlord & G. Low (eds.), New directions in language testing, 3 (12). Oxford: Pergamon Press.

- Delamont, S. (2004). <u>Ethnography and Participant Observation</u>. In C. Seal, G. Gobo, J. F. Gubrium & D. Silverman (eds.), Qualitative research practice, (pp. 205-512). London: Sage publications.

- Denscombe, M. (2010). <u>The Good Research Guide: For small-scale social research projects</u>. Mainden-Read: Open University Press.

- Denzin, N. K. (1978). <u>The Research Act: A theoretical introduction to sociological methods</u>. Eaglewood Cliffs, New Jersey: Prentice Hall.

- Dörnyei, Z. (2007). <u>Research Methods in Applied Linguistics</u>. Oxford: Oxford University press.

- Dunkin, M. J., & Biddle B. J. (1974). <u>The Study of Teaching</u>. New York: Holt, Reinhard and Winston.

- Ellis, R. (1994). <u>The Study of Second Language Acquisition</u>. Oxford: Oxford University Press.

- Enkstein, M A., & Noah, H. J. (1993). <u>Secondary School Examinations: Instrumental perspectives on policies and practices</u>. New Haven: Yale University Press.

- Ferman, I. (2004). <u>The Washback Effect of an EFL National Matriculation Test to Teaching and Learning</u>. In L. Cheng, Y. Watanabe; & A. Curtis (eds.), Washback in language testing: Research context and methods, (pp. 191-210). Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Fredericksen, N. (1984). <u>The Real Test Bias: Influences of testing on teaching and learning</u>, American Psychology, 39, 193-202.

- Fullan, M. G. (1983). <u>The New Meaning of Educational Change (1st ed.)</u>. New York: Teachers College Press.

- ……., ..., ... (2007). <u>The New Meaning of Educational Change</u>. New York: Teachers College Press.

- ……., …, ..., & Steigelbawer S. (1991). <u>The New Meaning of Educational Change</u>, London: Cassell Educational Limited.

- Geisinger K. F. (1994). <u>Cross-cultural Normative Assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments</u>. Psychological Assessment, 6, 304-312.

- Gibbons, P. (2006). <u>Bridging Discourses in the ESL Classroom</u>. London: Continuum.

- Gosa, C. M. C. (2006<u>). Investigating Washback: A case study using diaries</u>. Unpublished Doctoral thesis, University of Lancaster, England.

- Green, A. (2007). <u>IELTS Washback in Context: Preparation for academic writing in higher education</u>. Cambridge: Cambridge University Press.

- Guilham, B. (2000). <u>The Research Interview</u>. London: Continuum.

- Hammersely, M., & Atkinson, P. (1983). Ethnography: Principles on practice. London: Methuen.

- Hamp-Lyons, L. (1997). Washback, Impact, and Validity: Ethical concerns. Language Testing, 14, 295-303.

- Hawkey, R. (2006). Teacher and Learner Perceptions of Language Learning Activity. ELT journal, 60 (3).

- Henrichsen, L. E. (1989). Diffusion of Innovation in English Language Teaching: The ELEC effort in Japan, 1956-1968. New York: Greenwood Press.

- Hesse-Bibber, N. S., & Leavy, P. (2006). The Practice of Qualitative Research. Thousand oaks, California: Sage Publications.

- Hitchcock, G., & Hughes, D. (1995). Research and the Teacher. London: Routledge.

- Hsu, H. F. (2009). The Impact of Implementing English Proficiency Test as a Graduation Requirement at Taiwanese Universities of Technology. Unpublished Doctoral thesis, university of York, England.

- Hughes, A. (1993). Backwash and TOEFL 2000. Unpublished Manuscript, University of Reading, England.

- ……., ... (1988). Assessment. McMillan Hermann.

- Hymes, D. H. (1972). Functions of Language in the Classroom. New York: Teachers' College Press.

- Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). The Effects of Standardized Testing. London: Kluwen, Nijroff Publishing.

- Kennedy, C. (1990). Evaluation of the Management of the change in ELT Projects. Applied linguistics, 9 (4), 329-342.

- Kerlinger, F. N. (1970). Foundation of Behavioural Research. New York: Holt, Reinhart, and Winston.

- Khaniya, T. R. (1990). Examinations as Instruments for Educational Change: Investigating the washback effect of the Nepalese English examination. Unpublished Doctoral thesis, University of Edinburgh, England.

- Krashen, S. (1981). Second Language Acquisition and Second Language Learning. Oxford: Pergamon.

- Kvale, S. (1996). Interviews. London: Sage Publications.

- Ladico, M. G., Spaulding, D. T., & Voegtle, K. H. (2007). Methods in Educational Research: From theory to practice. San Francisco, California: Jossey Bass.

- Laing, R. D. (1967). The Politics of Experience and the Bird of Paradise. Harmoudsworth: Penguin

- Lam, H. P. (1993). Washback- Can It Be Quantified? A study on the impact of English examination in Hong Kong. Unpublished Master dissertation, university of Leeds, England.

- Li, X. (1990). How Powerful Can a Language Test Be? The MET in China. Journal in Multilingual and Multicultural Development, 15 (5), 393-404.

- Li, D. (2001). Teachers' Perceived Difficulties in Introducing the Communicative Approach in South Korea. In D. Hall& H. Herwings (eds.), Innovation in English language teaching (pp. 149-169). London: Routledge.

- Linn, R. L. (2000). Assessment and Accountability. Educational Researcher, 29 (1), 4-16.

- Madaus, G. F. (1988). The Influence of Testing on the Curriculum. In C. N. Tanner (ed.), Critical issues in curriculum: Eighty-seven yearbook of the national society for the study of education. Chicago: University of Chicago Press.

- Mallinowski, B. (1992). Argonauts of the Western Pacific. London: Routledge and Kegan.

- Markee, N. (1993). The Diffusion of Innovation in Language Teaching. Annual Review of Applied Linguistics, 13, 229-243.

- ……., …. (1997). <u>Managing Curricular Innovation in language Teaching</u>. In D. Hall & A. Herwings (eds.), Innovation in English language teaching, 118-126, London: Routledge.

- Marshall, C., & Rossman, G. B. (1989). <u>Designing Qualitative Research</u>. Thousand Oaks: Sage Publications.

- McFee, G. (1992). <u>Triangulation in Research: Two Confusions.</u> Educational Research, 34 (3), 215-19.

- McNamara, T. (2000). <u>Language Testing</u>. Oxford: Oxford University Press.

- Mehan, H. (1979). <u>Learning Lessons: Social organisation in the classroom</u>. Cambridge, MA: Harvard University Press.

- Menard, S. (2002). <u>Longitudinal Research</u>. Thousand Oaks: Sage Publications.

- Messick, S. (1989). <u>Validity</u>. In R. Linn (ed.), Educational measurement. New York: McMillan.

- ……., …. (1996). <u>Validity and Washback</u>. Language Testing, 13 (3), 240-256.

- Miles, M. B. (1964). <u>Educational Innovation 1-48.</u> New York: Teacher College Press.

- ……., … …, & Huberman, A. M. (1994). <u>Qualitative Data Analysis: An expanded sourcebook</u>. Thousand Oaks: Sage Publication.

- Mitchell, R., Parkinson, B., & Johnstone, R. (1981). <u>The Foreign Language Classroom: An observational study</u>. Stirling Educational Monograph. Stirling: University of Stirling.

- Morgan, D. L. (1988). <u>Focus Groups: The qualitative research</u>. Beverley Hills: Sage Publications.

- Morrison, K. R. B. (1993). <u>Planning and Accomplishing School-centred Evaluation</u>. North Folk: Peter Francis Publishers.

- Morse, M., & Richards, L. (2002). <u>Readme First for a User's Guide to Qualitative Methods</u>. Thousand Oaks: Sage Publications.

- Moskowitz, G. (1971). <u>Interaction Analysis- A New Modern Language for Supervisor</u>. Foreign Language Annuals, 5, 211-221.

- Nisbet, J., & Watt, Y. (1984). <u>Case Study</u>. In J. Bell, T. Bush, A. fox, Y. Goodey, & S. Goulding (eds.), Conducting small-scale investigations in educational management (pp. 79-92). London: Harper and Row.

- Oppeinheim, A. M. (1992). <u>Questionnaire Design, Interviewing and Attitude Measurement</u>. London: Printer Publications Ltd.

- Patton, M. P. (1987), <u>How to Use Qualitative Methods in Evaluation</u>. Newbury Park, California: Sage Publications.

- Pearson, I. (1988). <u>Tests as Levers to Change</u>. In D. Chamberlain, & R. J. Baugardner (eds.), ESP in the Classroom: Practice and evaluation (pp. 98-107). United Kingdom: Modern English Publications.

- Pierce, N. (1992). <u>Distinguishing the TOEFL Reading Test</u>. TESOL Quarterly, 26 (4), 665-689.

- Popham, W. J. (1988). <u>The Merits of Measurement-driven Instrument</u>. Phi Delta Kappa, 68, 679, 682.

- Qui, L. (2004). <u>The Intended Washback Effect of the National Matriculation English Test in China: Intentions and reality</u>. Beijing: Foreign Language Teaching Press.

- Rapley, T. (2004). <u>Interviews</u>. In C. Seal, G. Gode, J. F. Grubrium; & D. Silverman (eds.). Qualitative research practice (pp. 15-33). London: Sage Publications.

- Read, J., & Hayes, B. (2003). The <u>Impact of IELTS on Preparation for Academic Study in New Zealand</u>. IELTS English Language Testing System Research Reports, 4, 153-266.

- Richards, K. (2003). <u>Qualitative Inquiry in TESOL</u>. Basingstoke: Palgrave McMillan.

- Robson, C. (2002). <u>Real World Research: A research for social scientist and practitioner-researcher</u>. Oxford: Blackwell.

- Rogers, E. M. (1993). <u>The Diffusion of Innovation (3rd ed.)</u>. London: McMillan.

- ………, …. …. (1995). <u>The Diffusion of Innovation (4th ed.)</u>. New York: Free Press.

- ……., ..., ... (2003). The <u>Diffusion of Innovation</u> (5th ed.). New York: Free Press.

- ……., ..., ..., & Shoemaker, F. (1971). <u>Communication of Innovation: A cross-cultural approach</u>. New York: Free Press.

- Shepard, L. A. (1990). <u>Inflated Score Gains: Is the problem old norms or teaching the test</u>? Educational Measurement: Issues and practises, 9, 15-22.

- Shohamy, E. (1992). <u>Beyond the Proficiency Testing: A diagnostic feedback-testing model for assessing foreign language learning</u>. The Modern Journal, 13 (3), 288-317.

- ……., ... (1993). <u>The Power of Tests: The impact of language tests on teaching and learning</u>. Washington, DC: National Foreign Language Center of Occidental Papers.

- ……., ... (2001). <u>The Power of Test</u>. England: Pearson Edition limited.

- ……., ..., Donitsa-Schmidt, S., & Ferman, I. (1996). Test <u>Impact Revisited: Washback effect over time</u>. Language Testing, 13, 298-317.

- Smith, M. C. (1991). <u>Put the Test: The effects of external testing on teachers</u>. Educational Researcher, 20 (5), 8-11.

- Spada, N., & Fröhlic, M. (1995). <u>The Communicative Orientation of Language Teaching Observation Scheme (COLT)</u>. Sydney, Australia: Macquarie University, the national centre for English language Teaching and Research (NCELTR).

- Stern, H. (1983). <u>Fundamental Concepts of Language Teaching</u>. Oxford: Oxford University Press.

- Taris, T. (2000). <u>A Primer in Longitudinal Data Analysis</u>. Thousand Oaks: Sage Publications.

- Thomson, R., Plumridge, L., & Holland, I. (2003). <u>Longitudinal Qualitative Research: A developing methodology</u>. International journal of Social Research Methodology, 6 (3), 185-187.

- Tsagari, K. (2006). <u>Investigating The Washback Effect of a High-Stakes EFL Exam in a Greek Context: Participants' perceptions, material design, and classroom applications</u>. Unpublished Doctoral thesis, University of Lancaster, England.

- Tuckman, B. W. (1972). <u>Conducting Educational Research</u>. New York: Harcourt Brace Jovanovich.

- Tunner, C. (2001). <u>The Need for Impact Studies of L_2 Performance Testing and Rating: Identifying areas of potential consequences at all levels of testing cycle,</u> in Brown, A. et al (eds.) Experimenting with uncertainty: Essays in honour of Alan Davies; Studies in Language Testing (11), 138-149. Cambridge: Cambridge University Press.

- Vermon, P. F. (1956). <u>The Measurement of Abilities</u>. London: University of London Press.

- Wall, D. (1996). <u>Introducing New Tests into Traditional Systems: Insights from general education and from innovation theory.</u> Language Testing 13 (3), 334-354

- ……., ... (2005). <u>The Impact of High-Stakes Examination on Classroom Teaching: A case study using insights from testing and innovation theory.</u> Cambridge: Cambridge University Press.

- ……., ..., & Horâk, T. (2006). <u>The impact of Changes in the TOEFL Examination on Teaching and Learning in Central and Eastern Europe: Phase 1, the Baseline Study</u>. TOEFL Monograph Series. Princeton, New Jersey: Educational Testing Service.

- Watanabe, Y. (1996). <u>Does Grammar-Translation come from the Entrance Examination? Preliminary Finding s from Classroom-based Research.</u> Language Testing 13 (3), 319-333.

- ……., ... (1997). <u>The Washback Effects of the Japanese University Entrance Examination of English Classroom-based Research.</u> Unpublished PhD thesis. Lancaster: University of Lancaster, UK.

- ……., ... (2004). <u>Methodology in Washback Studies</u>. In L. Cheng, Y. Watanabe, & A. Curtis (eds.), Washback in language testing: research context and methods (pp. 19-36). Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Web, E. J., Campbell, D.T., Schwartz, R. D., & Sechrest, L. (1966). <u>Unobtrusive Measures: Non-reactive research in the social sciences.</u> Chicago: Rand McNally.

- Weir, C. (1990). <u>Communicative Language Testing</u>. London: Prentice Hall International Ltd.

- Wesdrop, H. (1982). <u>Backwash Effects of Language Testing in Primary and Secondary Education</u>. Journal of Applied Language Study, 1, 40-50.

- White, R. V. (1988). <u>The ELT Curriculum Change: Design, innovation and management.</u> Oxford: Blackwell

- ……., ..., ... (1993). <u>Innovation in Curriculum Planning and Program Development</u>. Annual Review of Applied Linguistics, 13, 244-259.

- Wilson, N., & Mclean, S. (1994). <u>Questionnaire Design: A practical introduction</u>. New Town Abbey: University of Ulster Press.

- Wiseman, S. (1961). <u>Examination on English Education</u>. Manchester: Manchester University Press.

Appendix 1

<u>Current Study</u>: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

<u>Title:</u> Consent Letter

Dear colleagues,

   I am conducting a research into teaching and testing EFL classes at the secondary school level as part of a project at the University of Mohamed Khider of Biskra, and as one aspect of that project, at this phase of research, I wish to conduct classroom observation. In the long term, I may ask to carry out an experimentation of an Alternative Testing Model with your classes.

   I am seeking you consent to make your classroom and your pupils one of my experimental classes; during the observation session written, and audio data will be gathered and later analysed and presented in a written form as part of the final research. No unwished information or pupil will be named or identified in this report.

   If you consent to your class being used as experimental group, please sign the attached consent form and return it to the researcher. You are entitled to withhold your consent for this project, and you may withdraw your consent at any time.

   When the research is complete, you my request a copy of any data relating specifically to you class.

   If you have any questions or comments relating to this project, you may contact me at the following address.

Yours sincerely

June 2011.

Hoadjli Ahmed Chaouki
Doctoral researcher
Tel.: 06.62.51.06.23
Email: toufikme2001@yahoo.fr

English Division
Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

I consent the participation of my class in the observations and research project being carried out by HOADJLI Ahmed Chaouki.

School name: ………

Teacher name: ……

Tel. number: ………

Email: ………

Date:_____

Teacher signature

Headmaster signature

Questionnaire Validation Form

I hereby certify that I have read the teacher and the students questionnaires in this study conducted by Ahmed Chaouki Hoadjli who is working on Phd from the English Division at the university of Biskra, Algeria. I have provided the researcher in the present study remarks and comments in relation to the contents and layout of this data collection method.

Background Information on the Expert:

Name: ……………………………………………………………..

Name of University: …………………………………………...

Present Occupation: …………………………………………...

Degree: …………………………………………………………

Telephone Number: ……………………………………………

Email Address: …………………………………………………

Signed: …………………………………………………………

Researcher Contacts Details:

Ahmed Chaouki Hoadjli

A Doctoral Researcher

English Division

Department of Foreign Languages

Faculty of Letters and Foreign Languages

University of Biskra

Mobile: (+213) 0662510623

Email: ac.hoadjli@univ-biskra.dz.

Appendix 3

<u>Current Study</u>: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

<u>Title</u>: A Teacher's Questionnaire (01)

Dear colleagues,

I am carrying out a research on teaching and testing English as a foreign language in the Algerian secondary school classes. I would like to ask for opinions on this particular subject. Please complete the present questionnaire based on your own experience. All information will be treated in the strictest confidence.

Thank you for your collaboration.

If possible, you can provide me with additional comments at this E-mail: toufikme2001@yahoo.fr

N.B: Achievement test are meant to test the degree of achievement of students in relation to the instructional objectives and the taught syllabus. (Trimestral tests/ Examens trimesteriels/ الاختبارات الفصلية).

Section 1:

Please tick the appropriate answer or provide written answers.

1. What do these testing terms mean to you?

☐ Test: …..………………………...………….......………..……….....………………………

………………………………………………………………………………………………….

☐ Achievement test: …..………………………………...…..……………………………………

……………………………………………………………………………………………….....

☐ Test design: …..…………………………………………………………………....…………

………………………………………………………………………….…………………………

☐ Test development: ..…………………………………………………….……………………...

……………………………………………………………………………………….……………

☐ Washback in language testing: ……………………………………….......……………………..

……………………………………………………………………………………………….….…

2. Do tests provide you with appropriate information about:

- The student's level.
- The teaching programme.
- The appropriateness of the syllabus.
- The effectiveness of the instructional objectives.
- All these.
- Others: ……………………………………………………………….……………..

Please justify your answer:

…………………………………………………………………………………….…………………

………………………………………..………………………………………………………………

………………………………………………………………………………………………………..

………………………………………………………………………………………………………..

3. When developing achievement tests, what parameters do you take into consideration?

....
....

Please justify your answer: ....

....

....

4. When developing achievement tests, do you consider your student's general language ability?

- Yes.

- No.

Please justify your answer: ....

....

....

5. What abilities do your achievement test target?

....
....

Please explain: ....

....

....

6. Before developing your achievement tests, what theoretical bases do you rely on?

....
....

*Please justify your answer: ....

....

....

....

7. Do you link your test-items in achievement tests to the content of the syllabus?

- Yes.

- No.

If yes, please say how? ………………………………………………………………….………..

………

………

………

8. Are you tests quantity-based as far as the syllabus is concerned?

   - Yes.

   - No.

9. Do you find scores often relevant to the students' achievement and progress in accordance with the instructional objectives?

- Yes.

- No.

If yes, please say how? …………………………………………..………………………………..

………

………

………

10. When developing achievement tests, do you consider the relationship between test-content and the instructional objectives?

- Yes.

- No.

If yes, please say how? …………………………………………..………………………………

………

………

………

11. When developing achievement tests, do you take into account your students'...............................?

- Personality characteristics.

- Background knowledge.

- Aptitude for learning.

- Native language.

- All these.

- Other: ……………………………………………………………………………………...

Please justify your answer: ……………………………...…………………………………………...

……….

……….

………

12. Do you think your achievement tests can serve your teaching?

- Yes.

- No.

Please justify your answer: ……………………………...…………………………………………

……….

……….

……….........

13. Do you take into consideration the affective side of your students when designing and developing achievement tests?

- Yes.

- No.

- Sometimes.

 Please justify your answer: ……………………………………………………………………………...

………...

………...

14. Do you think your students learn from achievement tests?

- Yes.

- No.

Please justify your answer: …………………..……………………………………………………………..

……..

……..

……..

15. Do you think that achievement tests concern other broader areas?

- Yes.

- No.

Please justify your answer: …………………………………..…………………………………………..

…….

…….

…….

16. Do you think you need to be assisted in order to develop achievement tests?

- Yes.

- No.

Please justify your answer: ……………………………………..…………………………………………..

…….

…….

…….

17. What kinds of resources are not available to carry out your achievement test?

  Please justify your answer: ..……………………………………………………………………………

…….

…….

…….

Section 2:

Please tick the appropriate answer or answers.

18. Always   Sometimes   Rarely   Never

Do you produce texts for your tests?
* In case you do not create texts, do you mention their source(s)?
* Do you provide your students with a glossary?
* Are your students familiar with the style of your tests?

19. What is/are the source(s) of your test texts?

- Textbook.
- Newspapers.
- Magazines.
- Internet.
- Other sources: (please specify)…………………………………………………………………………..
……

20. What parameter(s) do you consider in choosing your test-texts?

- Content of the programme.
- Level of students.
- Time of the test.
- All of these.
- Other parameter(s): (please specify)..……………………………………………………………......
……

21. Do you specify the vocabulary included in the test-text from the content of the syllabus?

- Yes.

- No.

- It depends.

 Please justify your answer: …………………………………..………………………………………………..
……
……
……

22. Do you test lexis ………………?

- In relation to the reading comprehension section.
- Independently from the text.
- In other ways.

Please justify your answer: …………………………………………………………………….……..

……

……

……

23. Do you test grammatical structures on the basis of:

- Structures taught in class.
- Structures included in the test-text.

Please justify your answer: ……………………………………………………...…………………………..

……

………..

……

24. Do you limit the number of grammatical structures to be tested?

- Yes.
- No.
- It depends.

Please justify your answer: ………………………………………………………………………………

……

……

……

25. When testing the grammatical structure, do you:

- keep the same structure in each test.
- vary the structures from one test to another.
- It depends.

Please justify your answer: …………………………………………………...……………………………..

……

………..

……

26. When testing grammatical structures, do you use the same testing technique(s)?

- Always.
- Rarely.
- Never.

Please justify your answer: ……………………………………………………………………………..

……

……

……

27. Do you choose the testing techniques because they are:

- Easy for students.
- Easy to be scored.
- Part of the "Bac" examinations activities.
- Helpful to provide you with the effectiveness of the teaching programme.
- Other reasons: …………………………………………….....………………………………………..

Please justify your answer: …………………………………………………….....…….………

………..

………..

………..

28. Do you test the listening skill?

- Yes.
- No.
- Sometimes.

 Please justify your answer: ..………………….....………………………….....………………………

…….

…….

…….

29. Do you test the speaking skill?

- Yes.
- No.
- Sometimes.

Please justify your answer: ………………………………………………………………..……………..

………

……….....

………

   30. Which skill do you think is possible to test within your context of teaching?

- The listening skill.
- The speaking skill.
- Both.

Please justify your answer: …………………………………………………………………………....

……..

……..

……..

   31. Do you proceed with the same scoring method?

- Yes.
- No.
- It depends.

Please justify your answer: …………………………………….……………………………………

………..

………..

………..

   32. Do you inform your students about the criteria of your scoring method?

- Yes.
- No.
- Sometimes.

Please justify your answer: ……………………………………………………………………...…..

……..

……..

……..

33. How does this affect your students?

………
………
………
………
………

Section 3:

34. In your opinion, do teachers in the Algerian Secondary Schools give importance to testing? Please explain:

………
………
………
………
………
………
………
………
………
………

35. In your opinion, is it possible to partially modify, or completely change, the current testing model, applied by the secondary school teachers?

Please explain:

………
………
………
………
………
………

36. What would you suggest as an alternative?

Please explain: ....

....

....

....

....

....

-End of Questionnaire-

Thank you for your help.

June2011.

Hoadjli Ahmed Chaouki
A Doctoral researcher
English Division

Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

Appendix 4

Current Study: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

Title: Student's Questionnaire 1

Dear Students,

I am carrying out a research on teaching English as a foreign Language, and testing in the Algerian secondary schools. I would like to ask you for your opinions on this subject. Please complete the present questionnaire based on your own experience. All information you provide in this questionnaire will be treated in the strict confidence.

Thank you for your collaboration.

Section 1:

1. Do achievement tests cover the content of what you have learnt?
- ☐ To a great extent.
- ☐ To some extent.
- ☐ To a very little extent.
- ☐ There is no correspondence between the two.

2. Are the texts of these tests based on the content of the units you studied in class?
- ☐ Yes.
- ☐ No.
- ☐ Not always.

3. Do you think the texts are ....?

- ☐ Too long.

- ☐ Too short.

- ☐ Neither long nor short.

4. Is the language of these texts?

- ☐ Very difficult.

- ☐ Not very difficult.

- ☐ Not very easy.

- ☐ Very easy.

5. Is the number of difficult words that these texts comprise?

- ☐ Very big.

- ☐ Not very big.

- ☐ A few words.

6. Do you think the content of these texts corresponds your interests?

- ☐ Yes.

- ☐ No.

- ☐ Sometimes.

7. Before you answer the questions, how many times do you read the text?

- ☐ One time.

- ☐ Two times.

- ☐ Three times.

- ☐ It depends on the type of the text.

8. Does your reading(s) of a text take?

- ☐ A long time.

- ☐ A short time.

- ☐ It depends.

9. Are the questions often ………………….?

- ☐ Very clear.

- ☐ Clear.

- ☐ Not clear.

10. Are the testing methods used in these tests usually ……………………?

- ☐ The same in every test.

- ☐ Vary slightly from one test to another.

- ☐ Vary completely from one test to another.

11. Is the distribution of test-items among the test parts?

- ☐ Equal.

- ☐ Not equal.

- ☐ It depends from one test to another.

12. Which test-items are difficult to answer?
- ☐ Comprehension.
- ☐ Grammar.
- ☐ Lexis.
- ☐ Written questions.
- ☐ All these.

13. Do you think these tests test…………………….?
- ☐ Your degree of comprehension.

- ☐ The amount of vocabulary you know.

- ☐ Your mastery of grammatical structures.

- ☐ Your ability to use English language.

- ☐ All these.

14. To have good results in these tests, you need to………………?
- ☐ Revise continually.
- ☐ Revise a short time before the day of the test.
- ☐ Do not revise at all.

15. Is the scoring method of each test known to you?

- ☐ Yes.

- ☐ No.

- ☐ Not always.

16. Does your mark in English usually reflect your progress and achievement in that foreign language?

- ☐ Yes.

- ☐ No.

- ☐ To some extent.

Section 2:

17. In your opinion, what is negative of these tests? ……………………………………………………………
…….
………..…
…….

18. In your opinion, what are some suggestions you can give to your teacher to make these tests more suitable?
…….
…….
…….
…….

19. Do you have any additional comment on the subject of testing?
…….
…….
…….
…….

-End of Questionnaire-

Thank you very much for your help.

June2012.

Hoadjli Ahmed Chaouki
A Doctoral researcher
English Division

Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التربية الوطنية

الديوان الوطني للامتحانات والمسابقات

دليل بناء اختبار

في امتحان البكالوريا

مادة اللغة الإنجليزية

نوفمبر 2010

بسم الله الرحمان الرحيم

عرف امتحان البكالوريا في السنوات الأخيرة تطورا ايجابيا على مختلف مستوياته و على الرغم من هذه التطورات التربوية و التنظيمية و المادية أهمية الجهود التي بذلت فإن الواقع يشير إلى وجود تفاوت من مؤسسة إلى أخرى في نوعية مواضيع الاختبارات التي تقـدم الى التلاميذ، و في مقاييس بنائها و معايير تقويمها و تحديد أهدافها، بل أن البعض منها يبتعـد ابتعادا كبيرا على ما يقتضيه التقويم في البكالوريا.

من ثمة كانت الحاجة ماسة لوضع دليل حول كيفية بناء الاختبارات فـي مختلـف اختبارات البكالوريا حسب كل مادة و هذا ما يجعل عملية التقويم سهلة و هادفة . لذا فإن دليـل بنـاء الاختبارات وثيقة منهجية يستعين بها الأستاذ على بناء الاختبارات وفق قواعد علمية صـحيحة يتمكن من خلالها من تقويم الأهداف المسطرة في البرامج الرسمية و كذا المهارات و القـدرات التي يعمل على أن يكتسبها تلاميذه زيادة على انها وثيقة تكوينية تساهم في تكوين الأستاذ على كيفية بناء الاختبارات و اكتساب القدرة على ذلك.

و هي بالنسبة للتلميذ تساعده على أن يتعود في قسمه على طريقة معينة من الاختبارات حتى لا يتفاجأ في امتحان شهادة البكالوريا بطريقة مختلفة عما تعود عليه في الثانوية بل يجد نفسه أمام وضع مألوف و منهجية مطروقة من قبل فيطمئن و تهدأ نفسه.

لذا فعلى كل أستاذ أن يسهر على بناء الاختبارات الفصلية وفق ما جاء في هذا الدليل و هذا ما يستوجب دراسـة هذه الوثيقة دراسة جادة، و تطبيق ما جاء فيها من منهجيـة فـي بنـاء الفروض و الاختبارات الفصلية التي ينظمها لتلاميذه في السنة الثالثة ثانوي حتى يتعودوا عليها و حتى يكتسبوا هذه القدرة و هذه المهارة فعلى الجميع دراسة هذه الوثيقة في جلسات خاصـة تنسيقية مع الأستاذ مسؤول المادة و بإشراف مدير المؤسسة في كل فصل من فصول السنة قبل موعد الاختبارات. حتى تكون الاختبارات المطروحة مبنية وفق هذا الدليل.

و على الأستاذ أن يحدد الأهداف المرجوة من كل اختبار، و يحدد كذلك القدرات و المهـارات الواجب تقويمها في الاختبارات المطروحة من طرفه، و عليه أن يقيم هذه الاختبارات من كـل الجوانب، خاصة الجوانب المعرفية و القدرات و المهارات وفق ما هو موجـود فـي البـرامج الرسمية المسطرة من وزارة التربية الوطنية.

و عليه فإن دراسة البرامج و تحديد المعارف و القدرات و المهارات أمـر ضـروري يعمـل الأستاذ على تحقيقه في القسم أثناء الدراسة، كما يجب عليه أن يعمل على تقويم هذه المكتسبات ومدى تحقيقه من خلال الاختبارات الفصلية.و إذا تحقق هذا ميدانيا نكون قد حققنا عدة أهـداف منها:

- بناء الاختبارات وفق قواعد علمية صحيحة محددة الغايات و الأهداف.
- التقويم الصحيح و السليم للأهداف و القدرات و المهارات المسطرة في البرامج الرسمية.
- توحيد طريقة الاختبارات على مستوى الوطن الرسمية منها أو غير الرسمية.
- تكوين الأستاذ على طريقة بناء الاختبارات و على تحديد الأهداف و المقاصد و أخيرا على تقويم هذه الأهداف و المقاصد.

- إعداد التلميذ لخوض امتحان البكالوريا بتعويده في الاختبارات الفصلية على نفس الطريقة التي تبنى بها مواضيع البكالوريا.

وفي الأخير و حتى نتمكن من تحقيق الأهداف المرجوة فإنني أنبه إلى ضرورة خضوع المواضيع المطروحة للمعايير التالية:

- أن تكون مبنية بطريقة تتماشى و طبيعة الاختبارات الموجودة في أدلة بناء الاختبارات المرسلة إليكم و الموجودة على مستوى المؤسسات التعليمية.
- التطابق التام مع البرامج الرسمية.
- التماشى و الوقت المحدد للاختبار حسب كل مادة و شعبة.
- أن يكون الموضوع في متناول المترشح المتوسط.
- تغطية أكبر نسبة من البرنامج الرسمي.
- أن لا تؤخذ من أي حولية أو أي امتحان سابق.
- تفادي كل الأخطاء سواء ما تعلق منها بالمادة أو اللغة.
- أن تصاغ أسئلتها بشكل واضح و دقيق و مفهوم و غير قابل للتأويل.
- أن تحرر بدقة ووضوح بخط مقروء.
- أن تكون قابلة للحل وفي الوقت المخصص لها.
- أن نتجنب المواضيع الطويلة جدا و التي تأخذ وقتا طويلا للقراءة من طرف المترشح.
- أن يعد مع كل موضوع إجابة نموذجية و سلم التنقيط.
- أن يسهر كل أستاذ على القيام بتصحيح اختبارات تلاميذه بعد الامتحان لتنبيههم إلى كيفية تحاشي الأخطاء الواردة أو المحتملة، وينور سبيلهم بكل ما يفيدهم لتجنب العثرات مستقبلا و يزودهم بمنهجية واضحة للإجابة عن أي سؤال يوجه إليهم.

و على الأستاذ أن يختم عمله بتقويم موضوع الاختبار وفق بطاقة التقويم التي زود بها هــذا الدليل.

وفقنا الله جميعا لما فيه خير للبلاد و العباد

مدير الديوان الوطني للامتحانات و المسابقات

INTRODUCTION

This document aims at supplying the teachers with some information and recommendations to help them prepare their pupils for the Baccalauréat English paper . It also provides useful guidelines for building Baccalauréat tests. The various types of tasks are introduced, explained , and wherever necessary, illustrated with examples.

Moreover, as far as the Competency Based Approach is concerned, it provides practitioners with a framework containing the disciplinary (target) competencies , namely interacting. (dialogue completion is a form of interaction) , interpreting , and producing

Users are invited, before setting a test, to consider the criteria listed in the preamble (Arabic version)

DURATION AND COEFFICIENT

Streams	Coefficient	DURATION
LANGUES ETRANGERES	05	03
LETTRES ET PHILOSOPHIE	03	02
Sc . Exp./ Math/ TM/ GE	02	02

1- RECOMMENDATIONS

Test the examinees only on what they have actually learnt during the academic year.

a-Congruency with

- official syllabus regarding topic, content and cognitive and learning objectives
- official instructions (types of activities, number of activities, timing , length)
- formulation of instructions : it is more convenient for the examinees if they are instructed in the same way they are used to in learning and testing during the academic year.

b- Test Elaboration

- **Secrecy:** Nobody should be informed of what text has been selected
- **Originality:** It should not have been given previously in any form
- **Correctness:** It should be mistake-free
- **Acceptance by all:** The text should not deliberately hurt or insult anyone
- **Meaning:** In spite of its size, the text should still be meaningful and coherent
- **Sources:** Full sources (author, title, publishing house, place , year) should be given
- **model answers**, including all sections
- **scoring scale** detailed and global
- It is recommended that:
- model answers and scoring scale be prepared beforehand by a different team , not necessarily the test setters, primarily to test the feasibility of the paper
- the activities should be graded from simple to complex

II) Objectives of the examination paper

The English examination paper aims to evaluate :

1. The candidates 'ability to understand and do tasks in connection with a reading passage, either adapted or authentic, based on a topic selected from the syllabus ,and to make good use of the language components.

2. The candidates 'ability to mobilize the appropriate resources to express themselves reasonably and correctly in a given situation (in a written form).

III)Organization of the examination paper.

IMPORTANT: It is advisable that the paper should revolve around the same theme, to be in accordance with the philosophy of the competency-based approach,i.e, tests provide examinees with opportunities to learn and re- use even while taking the test.

The examination paper is made up of the two following parts:

-Part One : Reading . This part consists of:

A. Comprehension and Interpretation.

This sub-part focuses on the global and detailed understanding of a reading passage through a number of comprehension-type tasks (see table page 4).

B .Text Exploration

This sub-part, through activities related to the reading passage or a similar topic, deals with the knowledge and use of the language(see pages 5,6,7,8).

It contains tasks bearing on:

1. vocabulary and morphology,

2. grammar,

3. sound system,

4. discourse.

Part two: Written Expression

This part consists of setting the testees to produce a piece of writing in a topic related to the text theme.

PARTS OF BAC PAPER	TYPOLOGY OF TASKS	TYPICAL INSTRUCTIONS	OBSERVATIONS	TEXT LENGTH & NUMBER OF TASKS		
PART ONE Reading Comprehension/ Interpretation LE-LP 07 points Sc-Exp-M-TM-GE 08 points		Read the text carefully and do the activities.	This instruction is to be put above the text.	L.E	L & philo	Sc.exp /M/ TM/G.E
	.choosing/giving the general idea of the text	.Choose/give the general idea of the text.	Giving(for L..E only)	Text of 200-250 Words	Text of 180-200 Words	Text of 120-180 Words
	.answering MCQ	.Write the letter which corresponds to the right answer.				
	.T/F statements	.are these statements true or false? write T or F next to the letter corresponding to the statement				
	.locating the right paragraph	.In which paragraph is it mentioned that...?				
	.Filling a table with information from the text.	.Fill in the following table.		5 to6 different activities	4 to5 different activities	4 to5 different activities
	.putting ideas in the right order according to their occurrence in the text.	.read the text and put the following sentences in the order they appear in the text.				
	.answering comprehension questions.	.Answer the following questions according to the text.				
	.choosing /giving a title to the text.	.Copy the title you think is the most appropriate/give a title to the text.	Giving(for L.E only)			
	.identifying type of discourse.	Circle the letter that corresponds to the right answer. The text is: a)descriptive, b)narrative...				
	.identifying type of text.	Circle the letter that corresponds to the right answer. the text is : a(a letter, a conversation)...				
	.Cohesive markers	.Who/what do the underlined words refer to in the text?				

B. Text Exploration L.E- 07 points L.P- 08 points Sc-Exp-M-TM-GE : 07points	.Matching words from the text with their definitions.	.match words and definitions.	Lexical items essential for the understanding of the text and those likely to be used in written expression should be included in the activities.	05 Activities 1from each: -voc1 -morpho1 -gram1 Sound sy1 Discour1	05 Activities 1from each: -voc1 morpho1 -gram1 Sound sy1. Discour1	05 Activities 1from each: -voc1 morpho1 -gram1 Sound sy1 Discour1
	.Finding words in the text whose definitions are given	.find in the text words whose definitions follow.				
1.Vocabulary and Morphology	.Matching words and their synonyms/opposites.	.Match words and their synonyms/opposites.				
	.Finding synonyms/opposites in the passage.	.Find in the text words that are synonyms/opposites to the following.				
	.Deriving nouns/adjectives/verbs from words in the text.	.Complete the chart as shown in the example.				
.	.Giving opposites to words in the text keeping the same root.	.Give the opposites of the following words keeping the same root.				
	.Dividing words into roots and affixes.	.classify the following words in the table.				

2.Grammar	combining statements with connectors provided.	Connect each pair of sentences with one of the words given. Make changes where necessary.				
	Asking questions to the underlined words.	Ask the questions which the underlined words answer.	The activities must cover a variety of grammar points and should not be limited to reported speech, conditionals or active/ passive voice.			
	Paraphrasing with prompts given.	Rewrite sentence B so that it means the same as sentence A.				
	Supplying the correct form(s) of the verbs in a given passage.	Give the correct form(s)of the verbs in brackets.	Give a short passage, not unrelated sentences.			

3.Sound system	Matching pairs that rhyme.	Match pairs that rhyme.				
	Classifying words according to final 'ed' or 's	Classify the following words according to the pronunciation of the final 'ed' or 's'.				
	Underlining the silent letter in a word.	Underline or write the silent letter in each of the following words				
	Stress pattern.	Classify the words according to their stressed syllable.				
	Number of syllables.	Classify the words according to the number of their syllables.				

4.Discourse	Filling gaps with a number of words.	Fill in the gaps with words from the list.	The activities suggested in Discourse should be conducive to the written expression.			
	Matching statements from column A with statements from column B.	Match statements from column A with statements from column B.				
	Re-ordering scrambled sentences into a passage.	Re-order the following sentences to make a coherent passage.				
	Dialogue completion.	Imagine what A/B says and complete the following dialogue.				

PART TWO Written Expression L.E-06points L.P-05pcints Sc-Exp-M-TM-GE 05 points	Choose ONE of the foilowing topics:	Topic one	One: Should be thematically related to the topic of the initial text.			
		Topic two	TWO: thematically related to other themes of the syllabus			

ADDENDA

A- CRITERIA TO SET A SITUATION OF INTEGRATION

1) A situation of integration requires :

-authentic documents

-well-formulated, readable (comprehensible) instructions rather than questions

2) A situation of integration should be:

- meaningful and motivating

-close to the learner's /testee's level

-adapted to the learner's/testee's level

-in accordance with the target competence(production)

-with in a situation of communication

-be supported by some hints /tips to guide the learner /testee

3) A target situation (a situation of integration) integrates linguistic resources to achieve a communicative purpose .

B- ASSESSING A SITUATION OF INTEGRATION

CRITERIA	INDICATORS
1.RELEVANCE	Learner's/testee's production aligns with requirements of a situation: *format*: a letter/ e-mail /invitation/instructions ... *objective*: to inform/ to describe/to complain... *targeted language*: functional language and vocabulary
2.SEMANTIC COHERENCE	Meaningful Logic in the information given Originality and development of ideas Organization : well-structured sentence &paragraph
3.SYNTACTIC COHERENCE AND CORRECT USE OF LINGUISTIC ELEMENTS	Correct use of grammar: tenses , word-order, subject-verb agreement Spelling, capitalization and punctuation
4.EXCELLENCE (creativity + vocabulary wealth)	New ideas , varied vocabulary items ,absence of repetitions and redundancy , legible handwriting, neatness.

NB: 1.Learners mustn't be penalized twice for the same mistake;

2.If only two of the first three criteria are met, the full mark is to be given.

3 - refer to BAC guide 2003 for further samples of tasks

School: ___ high school ___	School year :2011/2012.
Level: Second year FL ___	Duration :2h ___

THE SECOND TERM EXAM

Part One: (14 points).
Read the text carefully then do the activities.

Dream is the story *that* a person "watches" or appears to take part during sleep. Dream events are imaginary, but they are related to real experiences in the dreamer's life. They seem real to the dreamer while they are taking place. There are many types of dreams. Some are pleasant, others are annoying, and still others are frightening.

Everyone dreams, but some people never recall dreaming. *Others* remember only a little about a dream they had just before awakening and nothing about earlier dreams. No one recalls every dream and, in general, dreams are very easily forgotten.

Many experts who study dreams also feel that *they* are related to deep whishes and fears of the dreamer, and several theories explaining the meaning of dreams have been developed. Freud suggested that dreams are fulfillments of wishes, usually in disguised form. Some scientists have suggested that biological discoveries about dreams have made psychological theories of dreaming, such as freud's, unnecessary or false. These scientists argue that a dream is a meaningless response of the cerebral cortex to random stimulation from the brain stem.

The biology involved does not make dreams meaningless any more than it makes waking thought meaningless. Most psychologists and psychiatrist still consider dreams psychologically meaningful.

World book 2004

Reading and interpreting (7 PTS)

1 – Choose the most suitable title to the text (0.5pt)
    a – types of dreams   b – dream events   c – meaning of dreams

2 - Say whether the following statements are true or false according to the text. (02pt)
    a – dream events are real experiences in life.
    b – None can recall dreams
    c – Some dreams are frightening.
    d – Experts do not agree on the explanation of dreams.

3 - Answer the following questions according to the text. (2pts)
    a- how does the author define a dream?
    b- Do scientists agree with feud's suggestion about dreams?

4- What do the underlined words refer to in the text. (01.5pts)
    That (§1)-------------
    Others (§2)-------------
    They (§3)-------------

-How many theories explaining dreams are mentioned in the text? (1pt)

B / TEXT EXPLORATION (7PTS)

1 – Find in the text words closest in meaning to: (2pts)
    bothersome (§ 1) regard (§ 4)
  – Find in the text opposites to:
    real (§ 1)   exposed (§ 3)

2- Complete the following conditional sentences (2pts)
  a – **If** you ....(to dream) of a baby .You ....(to receive) Good news.
  b – The teacher may.............(to punish) you **if**...............(not do) the home work

3- Complete the chart below as shown in the example (1.5pt)

verb	noun	adjective
To help	help	helpful
....	dream	....
To mean	....	....
....	....	suggested

4- Reorder the words below to get a meaningful sentence(0.5pt)

  Are -that- thoughts - a - dreams - images- or - during - appear - sleep - in – person's - mind.
    Dreams..

5- Circle the stressed syllables in the following words. (1pt)
    Necessary - stimulation - psychology - realistic   •

Part two: Writing (06pts)

Choose one of the following topics:
Topic one:
    Dreams do indeed offer opportunities for fun, adventure, wish fulfillment, creativity, deep personal insight healing, they offer all this at no cost and with no line-ups!
    -write a short composition talking about the importance of dreams in people's daily lives.

Topic two:

  Imagine that your friend finds himself/herself in a dilemma. Suggest to him/her a solution to get out of it.
    - write a letter in which you suggest a solution to that problem.

FIRST TERM EXAM

Read the text carefully then do the activities

PART ONE : A/ Reading & interpreting (7pts)

Organization of Petroleum Exporting Countries (OPEC), international organization primarily concerned with coordinating the oil policies of *its* member states. OPEC was founded in September 1960. It has since been joined by Algeria (1969). Currently, the 11 member states supply over 40 per cent of the world's crude oil, and control nearly 80 per cent of known reserves. OPEC headquarters is in Vienna, Austria. Representatives of the member governments, meet at least twice a year to formulate policy.

In the late 1950's, world oil production considerably exceeded demand. The price of oil dropped, and OPEC was formed in reaction to this drop in payments. In the early 1970's international demand for oil began to exceed supply, and between 1973 and 1974 OPEC brought about a near fourfold increase in international crude petroleum prices to almost US $12 a barrel. In 1979 and 1980 OPEC members voted for a second round of price rises *which* pushed petroleum prices above US $30 per barrel and exacerbated inflation problems in the industrialized countries.

1. **Answer the following questions according to the text. (3pts)**

   a- What's the role of OPEC ?

   b- Why was OPEC formed ?

   c- Did it succeed in rising oil prices ?

2. **Say wether the following statements are True or False. Correct the false ones.(2pts)**

   a- OPEC headquarters is in Australia.

   b- OPEC was formed as a result of the fall in petroleum prices.

   c- The world's need for petroleum products is decreasing.

   d- Inflation is the natural result of oil price rises.

3. In which paragraph it is mentioned that **the rise in oil prices led to the rise in prices and fall in the purchasing value of money.(1pt)**

4. **What do the underlined words refer to in the text.(1pt)**

B/ Text Exploration (7pts)

1. **Find in the text words that are closest in meaning to the following.(2pts)**

   Natural (1§) = ...   fall = (2§) ....

2. **Find in the text words that are opposites to the following.**

   National ≠ (2§) ....   below ≠ (2§) ....

3. **Connect these pairs of sentences using the right relative pronouns and make any necessary changes.(who , which ,that, whose, where) (2pts)**

   a- OPEC coordinates the oil policies. OPEC has eleven members.

   b- OPEC is an organization. Its task is to formulate oil policies.

   c- Representatives of the OPEC members meet twice a year to formulate policies. They are eleven.

   d- Vienna is the capital of Austria. The OPEC headquarters is situated in vienna.

4. **Spot the mistake and correct it.(1.5pts)**

   Last weak , my teacher asked me to right a report about the OPEC.Eye asked my friend to help me ,bat he said know.Finally I managed too do it alone and the teacher was satisfied.

5. **Classify the following words according to the pronunciation of the final "s".(1.5pts)**

   governments –rises –policies – members – prices –states –

/ s /	/ z /	/ iz /

PART TWO :Written Expression.(6pts)

   Choose either topic 1 or topic 2.

   TOPIC ONE : Imagine you are on a **presidential election compaign** . Advertise your programme by writing a **policy statement** on how you plan to **fight** such **economic problems** like **inflation**.Use **going to** and the clues below.

   Reduce taxation – encourage private entreprise – encourage additional investment & higher productivity – renewed economic growth .

   Start like this :

   If I 'm elected to office, I**'m going to**...

   TOPIC TWO :summarize the text in your own words in no more than 6 lines.

2nd TERM EXAM OF ENGLISH

PART ONE: Reading   (15 poin
Read the text carefully then answer the questions:

Every year rivers carry amounts of waste into the sea. The sea also receives large quantities of chemicals from factories, smoke and gases from cars and even nuclear radioactive materials in some deep areas.

Ecologists are now wondering whether the ocean can be used as a waste basket for a long time. Throwing rubbish into the sea is considered as a forbidden act because it will harm people's health and destroy marine life and the beauty of coasts. Countries with miles of coastline must protect water from pollution. Nuclear dust has already killed a lot of fish and marine life, and sooner or later humans will be poisoned too.

Countries must agree to safeguard water surfaces. That sources which in fact is a place where human, animals, plants live with .Therefore, countries or nations must put many principles to stop that catastrophes for the purpose of living in a very clean, peaceful and good atmosphere such the water life.

A/ Comprehension/ Interpretation: (8points)

1-Are these statements _true or false? Write T or F next to the letter corresponding to the statement_.
   a. Every year rivers carry amounts of garbage into the sea.
   b. Humans will not be poisoned by nuclear dust.

2-Answer the questions according to the text.
   a- Why is throwing rubbish into the seas considered as a forbidden act?
   b- Has nuclear dust killed many kinds of fish and marine life?
   c- What should countries do to safeguard the water surfaces?

3-In which paragraph is it mentioned that:
   a- We mustn't through rubbish into seas since it is a illegal act
   b- Water surfaces are a place for animals and plants.

B/ Text Exploration: (7 points)

1.a-Find in the text the right synonyms of these words:
   -hold=.........(§1)   damage=........(§2)
   b-Find in the text the right opposites of these words:
   - some #....... (§3)   - die #.......... (§3)

2. Rewrite sentence 'B' so as it means the same as sentence 'A'.
   a- Pollution will destroy the livelihood of millions of animals.
   b-The livelihood of millions of animals.............................
   a- Human being must save the nature.
   b- The nature ....
3.Put the verbs between bracket into the correct tense:
   - We (to get) ill, if we (to drink) from this polluted water.
   -If you (to put) water at -100 degree , it (to freeze).

4. Mark the primary stress in the following words :
   -pollution / - sociology   /   - realistic   / - phonology

Part Two: Writing   (5pts)
Choose one of the following topics:

Topic one: You are an ecologist. The government asked you a program for fighting water pollution . Write your report.

Notes to use:
What must/ should be done :
  -to teach anti-pollution programs at schools
  - to make advertisements on TV
  - to put special budgets for cleaning the environment
  - to punish severely people/institutions causing pollution

Topic two: -Read the report written by a group of citizens.

The beautiful area we used to live in was spoiled: The rivers were polluted. The grass around the ponds was burnt, the trees were cut and all the beautiful flowers picked. The animals were killed and their habits destroyed.

You are the mayor of this city. Make a press release to apologise for the damage caused and to promise reparation.

** No Bees No Honey, No Work No Money**

Appendix 7

Script One :

I am happy to join you today in what will go down in history as the greatest demonstration for freedom in the history of our nation. For score years ago , a great American , in whose symbolic shadow we stand today , signed the emancipation Proclamation . This momentous decree came as a great beacon light of hope to millions of Negro slaves who had been seared in the flames of withering injustice . It came as a joyous day break to end the long night of their captivity . But one hundred years later , the negro still not free . One hundred years later , the life of the negro is still sadly crippled by the manacles of segregation and the chains of discrimination . One hundred years later , the negro lives on a lonely island of poverty in the midst of a vast ocean of material prosperity . One hundred years later , the negro is still languishing in the corners of American society and finds himself an exile in his homeland . So we have come to dramatize a shameful condition.

Script Two :

I say today , my friends , so even though we face the difficulties of today and tomorrow , I still have a dream . It is a dream deeply rooted in the American dream .

I have a dream that one day this nation will rise-up and live out the true meaning of its creed : " we could hold the truths to be self-evident that all men are created equal ".

I have a dream that one day on the red hills of Georgia the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood .

I have a dream that one day even the state of Mississipi, a state sweltering with the heat of injustice , sweltering with the heat of oppression , will be transformed into an oasis of freedom of justice .

I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character.

First Achievement Test
(Sub-Test One)

SECTION ONE : Listening .

Task One : **You are going to hear a speech . Listen to it and choose the right answer . Put a star (*) next to the right answer . (04 pts).**

01 – The person who is delivering this speech is :

    A – Nelson Mandela .
    B – Martin Luther King .
    C – President O'Bama .

02 – The speech is about :

    A – The rights of children to get education .
    B – The rights of women to get a job .
    C – The rights of Black people to get freedom .

03 – The person , who is delivering this speech , had celebrated a great event after ... :

    A – One hundred year .
    B – Two hundred years .
    C – Three hundred years .

04 – The person , who is delivering speech , is adressing :

    A – young people in America .
    B – Black people in America .
    C – All American people .

Task Two : **Listen to the speech again and say whether the statements are True (T) or False . Put (T) or (F) next to the chosen answer : (04 pts)**

    A – The speaker joined the demonstration to ask for equality and freedom of the black people in America . ...
    B – The speaker hoped that his children would be able to have the same rights as the children of the whites in America . ...
    C – The speaker wished that he would be judged on the content of his character and not on his colour . ...
    D – The speaker expected that he would be chosen the first black American president . ...

Task Three : Listen again to the speech and answer the questions below . (08 pts) .

A – Who was the orator (= speaker) ? .

....
....

B – What was he talking about ? .

....
....

C – What were his hopes and wishes ? .

....
....

D – The speech was oratorical . What are the characteristics of this style ? .

....
....

Task Four : In a few lines , reproduce (= summarize) what this speech was about . Did You enjoy listening to this speech ? say why or why not ? . (04 pts) .

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

- GOOD LUCK -

Secondary School : ....
Level : Second Year .
Class : ....

Date : ....
Full Name : ....

First Achievement Test
(Sub-Test Two)

SECTION TWO : Speaking (Written oral) .

Task One : <u>The underlined letters in the words below are mispronounced . Correct the mistakes in the transcriptions . (03 pts) .</u>

A – He put the <u>letter</u> / lɪ: tər/ in the mail box .(/   /).
B – Thousands of b<u>eau</u>tiful /buːtiful/ trees were burned . (/   /).
C – They w<u>ere</u> / wear/ saved . (/   /).
D – The rivers have been<u>po</u>llutted /pɔljuːtəl/ . (/   /).
E – Many b<u>ir</u>d / bird / species are in danger of extinction . (/   /).
F – The car has been wa<u>shed</u> /wɔʃpəl/ . (/   /).

Task Two : <u>Mark intonation with : ➚ and ➘ at the end of the sentence . (03 pts) .</u>

A – Will you come tomorrow ? ....
B – Where are you from ?......... .
C – Can you bring me your book please ?......... .
D – I wonder if you could explain to me the meaning of this word ?..........
E – Do you have a pen ?......... .
F – How are doing now ? ....

Task three : <u>Circle the silent letters in the words below . (03 pts) .</u>

A – Salt   B – Often
B – Know   C – Write
D – Cough   E – Could

Task Four : <u>Say how these words are pronounced . (03 pts) .</u>

A – Cut /   / .   B – Shirt /   / . C – Shirt /   / .
D – Two /   / .   E – Car /   / . F – Book /   / .

Secondary School : ....
Level : Second Year .
Class : ....

Date : ....
Full Name : ....

First Achievement Test
(Sub- Test Three)

SECTION ONE : Reading Comprehension (06 pts) .
Read the passage below , then , do answer the following tasks .

01- It is a great thing to go to jail for your right , but whether you are there for right or wrong , when you hear the big steel door close behind you and that they turn , you know you are in prison . That was Birmingham , May of 1963 . Martin luther king asked me to come down . I arrived at 11:30 a.m. on a Monday , and an hour and a half later , I went to jail with more than 800 other demonstrators . It was my first time to stay in jail .

" you Dick Gregory ? ".
" I'm Mr. Gregory ".

02 - Somebody snatched my collar and my feet did not touch the floor again until I was in solitary confinement , alone in my cell .

03 - Later in the afternoon I was brought downstairs and put in a cell built for twenty-five people There must have been 500 of us in there . When they moved us to eat , the corridors were so crowded that it was impossible to walk . You could not lie down and sleep .

04 - There was a little boy , may be four years old , standing in the corner of the cell , sucking his thumb . I felt sorry for him . He did not have even someone his age with whom he could play. I rubbed his head and asked him how he was

" All right " , he said .
" What are you here for ? " .
" Teedom " , said .

He could not even say freedom but he was in jail for it .

(Adapted from Teedom by Dick
Gregory , Nigger , BAC , 1993)

GLOSSARY :

Steel : Hard metal .
Come down : to come from one place to another .
Snatched : taken away .
Collar : the part around the neck of a chirt , jacket or coat .
Confinement : to stay in a close place .

Task Two : <u>Divide the following words into roots and affixes (prefixes + suffixes).
Use the table below .</u>

(a – organization   b – incapable   c – hopefully   d – international).

	Prefix	Root	Suffix
a			
b			
c			
d			

Task Three : <u>Fill in the blanks with one of the following : " cannot " / " can " / " will
be able " / " was able to " .</u>

a – The league of nations .... impose economic sanctions on warlike nations .

b – UN peacekeeping troops , called the ' Blue Hemlets ' , ... use force for temp-
orary self-defence purposes .

c – Germany .... to join securuty council soon because it is the thirld economic power in the
world .

d – The United Nations Educational , Scientific , and Cultural Organization (UNESCO) ....
launch its peace programme only after the end of the cold war .

Task four : <u>Spot the mistake then correct it .</u>

The united Nations replaced the league of Nations in 1945 , just after the end if the World War 2 .
Today , most of the world's independent nations is members , and each one have one vote in the
General Assembly . Every member can refers its disputes with its neighbours to the United Nations
Security Council . The UN Council is the body of the United Nations which can took poitical ,
economic , or military actions to settle international disputes .

<u>The Mistakes are :</u>

01 _   03 -

02 -   04 –

Task One : <u>Give a title to this passage</u>. The Title is ....

Task Two : <u>The four sentences below sum-up the main ideas of the passage . Skim through the text and match each sentence with its corresponding paragraph</u>.

A – Nobody is excepted from racism , even children were arrested . (.......).

B – Blacks were jailed in very bad conditions . (......) .

C – The jailor treated badly Gregory and all the other prisoners . (......).

D – Martin Luther King asked Gregory to participate in the demonstration . (........) .

Task Three : <u>Read again the passage , and say wheteher the following statements are true (T) or false (F)</u> .

A - Dick Gregory arrived in Birmingham early in the morning . (..........)

B – Dick Gregory was arrested at 13:00 a.m.   (...........)

C – Less than 500 people were arrested with Dick Gregory .   (...........)

D – The cell could not take 25 people .   (............)

SECTION Two : Language competence (08 pts) .

_Task One : <u>Match words in column A with synonyms in column B</u> .

A	B
A – Prison	1 – Liberty
B – Full of people	2 – Crowd
C- Solitary	3 – Alone
D – Freedom	4 – Jail

Task One : <u>Read the paragraph below , then , put the right punctuation marks and capitals</u> .

the horrors of modern warfare have made man think about ways of preserving human life in 1920 the league of nations was formed to keep international peace after the first world war the league of nations was able to resolve some conflicts around the world but it could not oppose the rise of totalitarianism in germany italy and japan hitler and mussolini were able to start the second world war because it had not enough power for its own

Task Two : <u>Use the following notes to summarize what happened to Dick Gregory</u> .

- received a call
- Birmingham , Monday
- arrived at 11:30
- met Martin Luther King
- marched in a big demonstration
- police arrived
- arrested with more than 800 demonstrators

The Audio-Text Script

Baby Plant : Mum , I'm hungry .

Mother : Then start photosynthesizing . The sun is shining bright use it .

Baby Plant : I know , I know , but my leaves need rest ... you know ...

Mother : Your leaves don't get tired .... They are meant to photosynthesize ... . Do you know how important we are for this world to go on ? ... let me explain .... :

Lives are very important to us ... all food people eat comes directly or indiectly from plants .... For example , apples come from an apple tree ... the flour used to make bread comes from a wheat plant .... Meat comes from a goat ... , and we all know that goats are animals , not plants , right ... , but, what does the goat eat ? .... It eats grass and hey ..., plants ...So , all the foods we eat come from plants . But , what do plants eat ? . They make their own food . Plants get their own raw materials for photosynthesis .

... from the soil , the air , and the sun . When it rains , water gets absorbed in the soil . The plants take in the water from the soil through its roots . In roots , there are tubes which are called : the xylem and phloem . The xylem lies towards the interior part of the root , whereas phloem lies towards the exterior part of the root . These tubes help in bringing water and minerals from the soil to leaves . Leaves breathe in carbon dioxide from tiny pores which are called ' stomata ' , which are on the under side of the leaf . Now , we have carbon dioxide and water , we are ready for photosynthesis , right ??? wrong !!!....

All we know are missing the most important ingredient ... the sun ... . Now , the process will start . It is in the leaf of the plant that most of the action takes place . If we cut the leaf and see it under a microscope , we see that right below under the outer layer of cells ...there is layer of soft elongated cells , called ' mysophyll cells ' ... inside the ' mysophylll cells ' are tiny notes which are called ' chloroplasts ' .... Chloroplasts contain a special chemical called chlorophyll ..... . It is chlorophyll which colours the leaves green .... When the light energy from the sun is captured by the chloroplasts , it is stored in stacks , called a ' granum ' , in which ' thylakoid membrane ' .... In these membranes , the chemical reaction of photosynthesis takes place so , remember : carbon dioxide ... plus water .... Plus sun ... gives you glucose ... plus oxygen .... The glucose or sugar is the food which the plant has made for itself and it is distributed to all parts of the body and oxygen is a waste gas which the plants breathe out ....people and animals use this very oxygen to breathe and continue living .....

Secondary School Name : ....
Level : ....
Class :............................ .

Date : ....
Full Name: ....

SECOND ACHIEVEMENT TEST
(SUB-TEST ONE)

SECTION ONE : LISTENING .

Task One : *You are going to hear an audio-text . Listen carefully to it , then , choose the right answer .*
Put a star (*) next to it .

01 – The audio-text is about .... :

    a- Green plants .
    b- The importance of oxygen .
    c- Photosynthesis .

02- Chlorophyll is contained in ... :

    a- Plant -leaves .
    b- The soil .
    c- Plant roots .

03- Photosynthesis takes place in ... :

    a- Human body .
    b- Green plants .
    c- Laboratory .

Task Two : *Say whether the following statements are true (T) or false (F) . Put (T) or (F) next to each*
statement .

    a- Photosynthesis is man-made process . (.......).

    b- CO2 is released during photosynthesis .(.......).

    c- Sun energy is not an element for the photosynthesis process . (........).

    d- All food people eat comes directly or indirectly from plants . (......).

Three : *Listen again to the audio-text and answer the questions below* :

    a- What is the process described ?

....

....

....

    b- What do plants need to make their own food ?

....

    c- Where is the sun light energy stored ?

....

    d- What colours the leaves green ?

....

....

....

our : *In few lines , reproduce (= summarize) the process described in the audio-text .*

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

....

SECOND ACHIEVEMENT TEST
(SUB-TEST TWO)

SECTION TWO : SPEAKING (WRITTEN ORAL) .

Task One : *Classify the following words according to their final 'ed' pronunciation .*

(founded – considered – stated – asked – produced – contained).

/t/	/id/	/d/

Task Two : *Circle the correct pronunciation .*

A) – Words

a – Plants **shoudn't be** cut down .

b – Forests **must be** protected .

c – Plants **can't be** replaced by man-made converters .

d – We **mustn't** hurt animals in danger of extinction .

B) – Pronunciation .

/ ʃudm bi: / . / ʃudnt bi: /

/ məst bi: / . / məs bi: / .

/ kənt bi: / . / kə mt bi: / .

/ məstnt / . / məsnt / .

<u>Task Three</u> : *Change the following transcriptions into their correspondent words.*

a - / ˈflðuð / : ....

b - / pðluːʃʌ / : ....

c - / weist / : ....

d - / zuː / : ....

e - / ʌeitʃð / : ....

f - / kweik / : ....

g - / dʒːɛi / : ....

h - / ʌoiz / : ....

<u>Task Four</u> : *Mark the intonation at the end of each sentence in the dialogue below (↗ or ↘).*

Car owner : *what should I do if the engine fails to start ?*

Mechanic : *check if there is fuel in the tank ?*

Car owner : *And if there is fuel in the tank .*

Mechanic : *then you should check if the battery is alright .*

SECOND ACHIEVEMENT TEST
(SUB-TEST THREE)

SECTION ONE : READING COMPREHENSION .
Read the passage below , then , do the following tasks .

01 – Noise pollution is a type of energy pollution in which distracting , irritating , or damaging sounds are freely audible . Noise pollution contaminants are waves that interfere with naturally-occuring waves of a similar type in the same environment . Sounds are considered noise pollution if they harmfully affect worldlife , human activity , or are capable of damaging physical structures on a regular , repeating basis .

02 – Noise pollution is chiefly created by unwanted sound like loud music , vehicles , aircraft , and rail transport . With houses built next factories , people experience noise pollution and its harmful effects .

03 – Noise pollution disturbs our health in a number of ways . Physiologically , apart from hearing troubles like deafness , noise pollution increases the rate of heat-beat , cholesterol , blood pressure and it causes headache . Psychologically , noise plollution leads to stress reaction like fear , irritatibility , and feeling of fatigue . As a matter of fact , increased noise levels cause a lack of concentration and reduces one's productivity and performance .

04 – Many noise problems can be prevented by avoiding living or working next to major crossroads , factories and shopping centres . Also build your home as far as possible from airports , as you should be a good neighbour by not disturbing those who live next door with music .

(Adapted from the Environment Encyclopedia , 2009)

Glossary :

- Contaminant : impure substance .
- Audible : that can be heard .
- Troubles : problems .
- Disturb : interupt .

Task One : *Choose the title that best suits the text . Put a star (*) next to it .*

a – Air Pollution .

b – Water Pollution .

c – Noise Pollution

<u>**Task Two**</u> : *Read again the text and complete the following table .*

- Causes of noise pollution	- - -
- Effects of noise pollution	- - -
- Preventing noise pollution	- - -

<u>**Task Three**</u> : *Read the third paragraph and complete the following diagramme .*

Noise Pollution

-
-
-
-
-

-
-
-
-
-

•
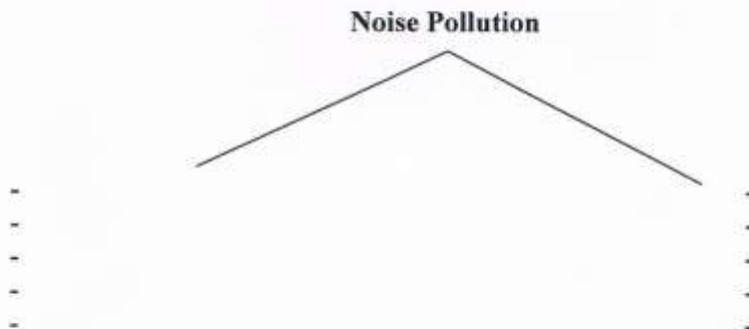
<u>**SECTION TWO : LANGUAGE COMPETENCE .**</u>
<u>**Task One**</u> : *Match words in column A with opposites in column B .*

A	B
a – near b – increase c – major d – harmful	1 – benificial 2 – minor 3 – far 4 – decrease

Task Two : *Complete the blanks in the table below with adjectives of your own . The first item is given to You as an example .*

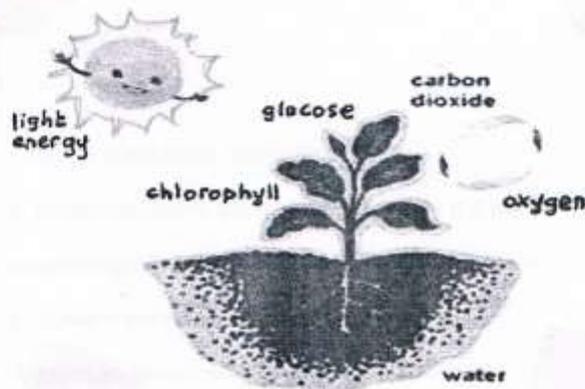| Suffixes	Meaning	Adjectives
able a- like b- ly c- some d- an	- that can be - resembling -regular of occurrence - which causes - nationality	- reliable -..................... -..................... -..................... -.....................

Task Three :*look at the picture below and fill in the blanks in sentences (a-d) using correct form of verbs .*

(- absorb – release – produce – capture)

a – Carbon dioxide (CO2) ....   by the stomata .

b – Water .... from the soil .

c – Oxygen (O2) ....   By the plant .

d – Sugar (glucose) ....

<u>Task Four</u> : *Spot the mistake and correct it .*

There are five types of alternative energies : solar , wind , biomass , and hydroelectric energies . All of them can be produce in our country because we have the following natural resource : the wind , the mountains However , solar energy is by far the more viable . It is cheaper , safer , and less polluting . Somebody has already build power plants in Biskra , Djelfa , Hasssi R'mel and other localities in the south of Algeria .

Their The mistakes		**Correction**	
a -	....	-	....
b -	....	-	....
c -	....	-	....
d -	....	-	....

SECTION THREE : WRITING .

<u>Task One</u> : *Read the paragraph below, then punctuate and capitialize it when it is neccessary .*

Without water life could not exist because water is essential to all life we must manage it intelligently and Carefully we think there is plenty of it in oceans lakes rivers and streams it may even seem as though we had More water than we need

<u>Task Two</u> : *Look at the picture below, then write a short paragraph about plant respiration .* •
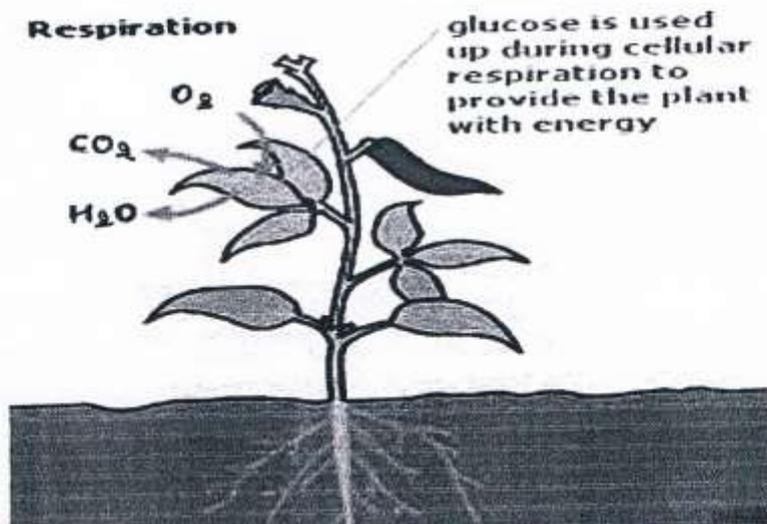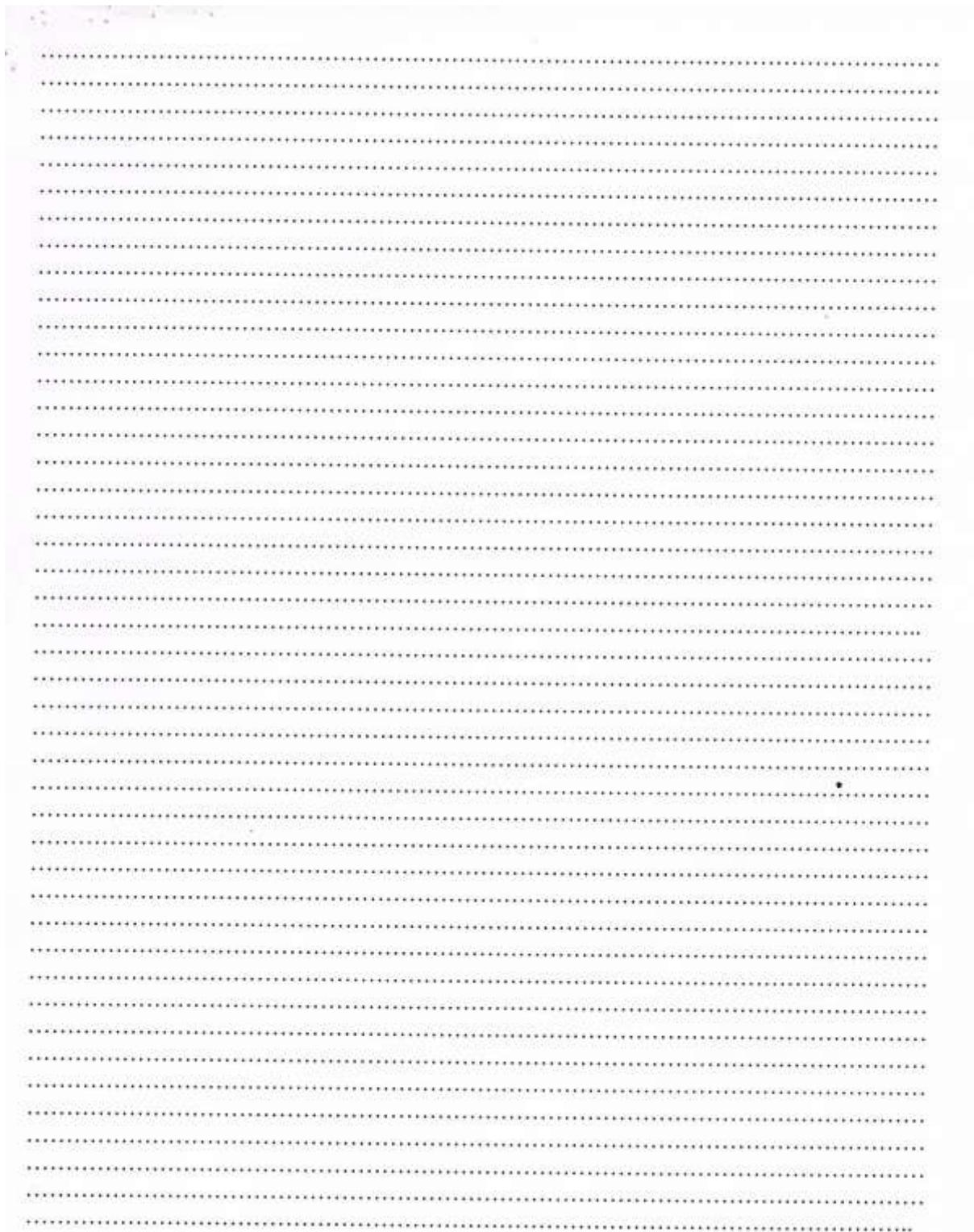
Respiration

O_2

CO_2

H_2O

glucose is used up during cellular respiration to provide the plant with energy

Current Study: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

Title: Teacher's Questionnaire 2

Dear colleagues,

The present questionnaire is part of a research on the washback effect of an Alternative Testing Model (ATM) on teaching and learning for EFL classes in the secondary schools in the Biskra region. Please complete this questionnaire based on your experience. All information will be treated in the strictest confidence.

Thank you.

Section 1: Demographic Information.

Please tick one appropriate answer or provide written answers.

1. Gender:

	Male	Female
Valid Percentage		

2. Your Age:

	21-30	31-40	Above 40
Valid Percentage			

3. Your highest academic level:

	'Bac' Certificate	License	Magister
Valid Percentage			

4. Number of years you have been teaching at the secondary school level:

	1-3	**4-6**	**7-9**	**Others**
Valid Percentage				

5. Level you currently teach. Tick more than one box, if necessary:

	Level 02	**Level 01+L03**	**Level 02+L3**
Valid Percentage			

6. Number of weekly sessions you teach for English classes:

	1-3	**4-7**	**Others**
Valid Percentage			

7. Number of hours you teach English per week in your secondary school:

	1-3	**4-7**	**Others**
Valid Percentage			

Section 2: Teachers' Perceptions of the Alternative Testing Model.

Please circle the appropriate answer.

8. What do you see as the major aims of implementing (trying-out) the Alternative Testing Model (ATM)?

		Not a major concern at all	**Of little concern**	**Of great concern**	**Main concern**
a-	**To prepare the pupils for their future career.**				
b-	**To cope with the present decline in English standards.**				
c-	**To motivate pupils to use integrated skills.**				
d-	**To direct pupils' learning.**				
e-	**To enable pupils to communicate appropriately outside the classroom.**				
f-	**To meet the instructional objectives set out in the taught syllabus.**				
g-	**To assess teachers through pupils' performance in the achievement tests drawn out from the ATM.**				

9. To what extent have you introduced the following activities in your teaching since the introduction of the Alternative Testing Model?

	Not at all	Some extent	A reasonable extent	A very large extent
a- Employ a new teaching method.				
b- Use a more communicative approach in teaching.				
c- Put more stress on individual performance.				
d- Put more stress on co-operative learning.				
e- Put more stress on group discussion.				
f- Put more stress on speaking.				
g- Put more stress on reading.				
h- Put more stress on writing.				
i- Put more stress on listening.				
j- Put more emphasis on the integration of the four skills.				
k- Encourage more pupil participation.				
l- Employ more real-life language tasks.				
m- Teach according to the achievement test format.				
n- Teach according to the 'Bac' examination format.				

Section 3: Washback on Teaching Materials.

Please circle the appropriate answer.

10. How do you often use the following resources in your teaching?

	Seldom or never	Sometimes	Most classes	Every class
a- Textbooks.				
b- Pictures or cards.				
c- Data-show (PowerPoint)				
d- Magazines.				
e- Newspapers.				
f- Examination guide.				
g- Language laboratory.				
h- Commercial books.				

11. What kind of extra-work or pressure if any do you think the Alternative Testing Model put on you in your teaching?

	No real increase	Little increase	Some increase	Large increase
a- Doing more lesson preparation.				
b- Preparing more materials for pupils.				
c- Revising the existing materials.				
d- Adopting new teaching methods.				
e- Setting-up new teaching techniques.				
f- Organizing more examination practice.				

Section 4: Washback on Classroom Teaching Behaviours.

Please tick one appropriate answer.

12. What is the medium of instruction you use when you teach in the classroom?

	English only	English supplemented with occasional Arabic explanation	Half English and half Arabic.	Mainly Arabic
Valid Percentage				

13. .How often do you do this in class?

	Almost never happens	Uncommon	Common	Almost the whole lesson
a- Talk to the whole class.				
b- Talk to groups of students.				
c- Talk to individual students.				

14. How often do you do the following activities in class?

	Seldom or never	Sometimes	Most classes	Every classes
a- **Explain the meaning of the text.**				
b- **Explain specific language items such as words or sentences.**				
c- **Go through textbook exercises.**				
d- **Organize pair work or group discussions.**				
e- **Organize integrated language skill tasks.**				
f- **Talk about achievement test.**				

Thank you for your co-operation

June 2012.

Hoadjli Ahmed Chaouki
A Doctoral researcher
English Division

Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

Appendix 9

Current Study: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

Title: Student's Questionnaire 2

Dear Students,

I am conducting a doctorate research at the University of Mohamed Khider of Biskra. I would like to ask you for your opinions on the proposed achievement test based on the Alternative Testing Model (ATM) by and through which you were assessed and evaluated during the last school year. Please complete this questionnaire based on your experience. All information will be anonymous and treated in the strictest confidence.

Thank you for your collaboration.

NB: The Alternative Testing Model (ATM) is the testing model on which achievement tests were drawn to assess your achievement in English language in relation to the instructional objectives set out in your syllabus. It includes three SUB-TESTS: SUB-TEST ONE (Testing Listening), SUB-TEST TWO: (Testing Speaking), and SUB-TEST THREE (Testing Reading and Writing and Language Components).

Section 1: Demographic Information:

Please tick the appropriate answer or provide written answers.

1. Gender:

	Male	Female
Valid Percentage		

2. Level:

	1st year	2nd year	3rd year
Valid Percentage			

3. Branch you study in:

	Natural Sciences	Letters and foreign languages	Letters and philosophy
Valid Percentage			

4. How many hours do you undertake to study English at school per week?

	1-4	4-9	More than 10
Valid Percentage			

5. How many hours do you undertake for the preparation of English lessons at home per week?

	1-3	4-7	More than 7
Valid Percentage			

Section 2: Washback Students' Attitudes Towards Teaching and Learning Behaviours

Please circle the appropriate answer.

6. How often does your teacher do the following in your English class?

	Usually	Sometimes	Rarely	Never
a. Talk to the whole pupils.				
b. Talk to groups of pupils.				
c. Talk to individual pupils.				

7. How often does your teacher do the following activities in your English class?

		Usually	Sometimes	Rarely	Never
a.	Tell your class the objectives of each lesson.				
b.	Demonstrate how to do particular language activities.				
c.	Explain specific language items, such as words or sentence structures.				
d.	Explain specific language structures.				
e.	Go through textbook exercises.				
f.	Assist pupils to answer exercises.				
g.	Talk about achievement tests.				

8. How often do you do the following in English class?

		Usually	Sometimes	Rarely	Never
a.	Doing pair work.				
b.	Doing group work.				
c.	Expressing your own ideas in English.				
d.	Asking questions.				

9. How often do you do the following in English outside class?

		Usually	Sometimes	Rarely	Never
e.	Talk to your teacher in English.				
f.	Talk to your classmates in English.				
g.	Watch TV programmes in English.				
h.	Read newspapers and magazines in English				
i.	Chat on the net in English.				

Please, tick the appropriate answer.

10. Do you agree with the following reasons for learning English?

	Agree	Disagree
j. **To get a better job**		
k. **To be able to study in higher education**		
l. **To be able to communicate with foreigners**		
m. **To have more and better opportunity in the future**		
n. **To satisfy my parents' desire**		
o. **To satisfy school requirements**		

11. Do you like to use the following strategies for learning English?

	Like	Dislike
a. **Reading newspapers and magazines in English**		
b. **Listening to radio programme in English**		
c. **Taking part in group activities in class in English**		
d. **Expressing opinions in English in class**		
e. **Putting more emphasis on understanding English**		
f. **Doing exercises and homework**		

12. Do you agree with the following opinions?

	Agree	Disagree
a. **Pupils like useful achievement tests**		
b. **Achievement tests are a valuable learning experience**		
c. **Achievement tests force pupils to study harder**		
d. **A pupil's score on achievement tests is a good indication of how s/he will be able to apply what has been learnt**		
e. **Achievement tests are one of the motivations for the pupils learning**		

Thank you for your co-operation

June 2012

Hoadjli Ahmed Chaouki
A Doctoral researcher
English division

Department of Foreign languages
Faculty of Letters and Foreign languages
University of Mohamed Khider, Biskra

<u>Current Study</u>: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

<u>Title</u>: Interview with Teachers

Mr. Hoadjli Ahmed Chaouki

Doctoral researcher, English division

The Questions

1. Would you introduce yourself?

2. What do these testing terms mean to you …..?
   a. Test.
   b. Test design.
   c. Test development.
   d. Achievement tests.

3. Before developing achievement tests, do you rely on any theoretical background?

4. Do you link the test –items of these tests to the contents of the syllabus on which you rest in teaching this foreign language?

5. Do you follow some precise stages/steps when developing your achievement tests?

6. Do you consider some psychological parameters of your pupils when developing your tests?

7. Do you think the scores your pupils usually obtain in these tests reflect actually their level?

8. Do you refer to the Baccalauréate Examination model to develop your achievement tests?

9. Do you think that the Baccalauréate Examination exerts any influence on you when developing achievement tests?

10. If yes, how?

11. Do you test the listening skill?

12. Do you test the speaking skill?

13. Do you think it is possible to test these two skills?

14. In your opinion, do you think the subject of testing is well considered in the teaching-learning process?

15. Do you have any additional comments on the subject of testing at the secondary school level?

-End of Interview-

Thank you for your help.

Hoadjli Ahmed Chaouki
A Doctoral researcher
English Division

Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

Appendix 11

Current Study: The Washback Effect of an Alternative Testing Model on Teaching and Learning: An exploratory study on EFL Secondary classes in Biskra.

Title: Interview and focus groups with students

Mr. Hoadjli Ahmed Chaouki

Doctoral researcher, English division

The Questions

1. Would you introduce yourself?

2. What does a test mean to you? And what's about an achievement test?

3. How do you feel during the period of tests?

4. Do you like or dislike tests? Why do you think pupils dislike tests?

5. Do your teacher of English language talk to you about the test before you do it?

6. How do you see (perceive) achievement tests? Do you think these tests correspond to what you have been taught in class? Or do you think the contents of these tests is relevant to the content of lessons?

7. Do the contents of these tests correspond to your interests?

8. Are the tasks/activities in these tests usually the same? Or do they differ from one test to another?

9. What about the distribution of test-items in the sections, do you think this distribution is balanced? Or do you think one section is given more importance than the others?

10. And what about the scoring scale, do you think it is clear enough for you? And does your teacher usually tell you about it (i.e., how he/she scores the activities and the sections?).

11. Now, Do you think the mark you usually obtain in achievement tests reflects your actual level in that foreign language?

12. Do you think that achievement tests reflect their attainment in English language?

13. In your opinion, what is good in these tests? And;

14. What is bad in them?

-End of Interview-

Thank you for your collaboration.

Hoadjli Ahmed Chaouki
A Doctoral researcher
English Division

Department of Foreign Languages
Faculty of Letters and Foreign Languages
University of Mohamed Khider, Biskra

Glossary

Achievement Test: It is used to measure the extent of learning in a problem prescribed domain often in accordance with explicitly stated objectives of a learning programme. The purpose of achievement tests is to discover how far students have achieved the objectives of course of a study.

Aptitude Test: It is used to measure the suitability of a test taker for a specific programme of instruction of a particular kind of employment.

Assessment: It refers to the process of quantifying the characteristics of a person.

Authenticity: It is the degree of correspondence of the characteristics of a given language test task to the features of a T.L.U. (Target Language Use).

Diagnostic Test: The primary purpose of diagnostic tests is to check out the strengths and weaknesses of learners. It identifies what they know and what they do not know.

Direct Test: It is precise, clear and straightforward. It requires test takers to perform in a precise way the skill we wish to measure.

Discrete-point Test: It is designed on the basis of measuring knowledge performance on a very limited scope.

Evaluation: It is a natural activity that can be formal or informal. It is something that may not always be made explicit, but may actually be undertaken unconsciously.

Examination: It is more general than a test since it embraces a combination of skills and proficiencies, and it is not related to a particular or to a precise area of learning.

Impact: The impact of a test is not restricted to the test itself alone, but it can also affect the other areas that are closely or loosely related to the test.

Indirect Test: It measures the abilities which underlie the skills we are interested in.

Integrative Test: It is to tap a greater variety of language elements in the completion of one task.

Interactiveness: It is determined by considering the characteristics of the test takers in relation to the tasks characteristics.

Objective Test: It is scored by comparing the test takers' responses with an established set of acceptable responses or scoring key.

Placement Test: It intends to gather enough information about test takers in order to place them at the appropriate stage of the teaching programme.

Practicality: It is concerned particularly with the ways to implement a test. A test which is said to be practical is a test which requires some necessary resources.

Proficiency Test: It measures language proficiency regardless of any language course, instruction, or previously experienced course.

Quiz: It is different from a test or an examination since its time is too brief and its focus is on a very narrow part of the context.

Reliability: It is about precise and repeatable measurement on clear bases. It is the consistency of the test's judgement and results.

Subjective Test: It is to require scoring by opinion and judgement.

Testing: It is a method in the sense it includes techniques, procedures and activities, which constitute an instrument of some sort, and that method needs an activity on the part of all the participants in the teaching operation.

Validity: It is attributed to the degree to which a test actually measures what is supposed to measure.

Washback: It refers to the influence tests have on teaching and learning.

ملخــص المذكــرة

تمحورت هذه الدراسة حول موضوع التقويم بصفة عامة، والتقييم بصفة خاصة للمنظومة التربوية في الجزائر، وبصيغة ادق دراسة شاملة لمنظومة الاختبارات الفصلية لمادة اللغة الإنجليزية في الطور الثانوي، وكيفية الوصول الى مطابقتها مع الأهداف التربوية المسطرة في المناهج التربوية.

تهدف هذه الدراسة الى تمكين أساتذة التعليم الثانوي من تطبيق الجانب النضري حول عملية التقييم وهذا ببلوغ الأهداف الرامية منه، حيث تصبح عملية التقويم عملية تربوية منظمة وليست فقط عملية روتينية الهدف من وراءها إعطاء علامة او ملاحظة للمتعلم.

فالدافع من وراء هذه الدراسة، قلة الاعمال والابحاث حول موضوع التقويم التربوي للغة الإنجليزية في التعليم الثانوي بالجزائر من جهة، ومن جهة أخرى ضعف عملية التقويم وقصرها والتي تظهر جليا في عدم تطابق المستوى الحقيقي للمتعلمين مع النتائج المحصل عليها في نهاية كل فصل دراسي. إضافة الى ذلك، أن عملية التقويم هذه لا تستند الى أسس علمية (نظرية أو إجرائية)، بل الى العشوائية والاعتماد الكلي على نماذج الامتحانات الرسمية التي تؤخذ كنموذج وحيد في عملية التقويم دون المراعاة الى وجوب أن تمس هذه العملية مختلف المهارات والقدرات عند المتعلم.

وكحل لهذه الوضعية، اقترحنا في هذه الدراسة نموذج (ATM) بحيث يمكن للأستاذ أن يبني اختبارات فصلية تعكس المستوى الحقيقي للمتعلم وهذا بعد اخذنا لكل النقائص الموجودة في الاختبارات التقليدية.

ولتطبيق هذا النموذج (ATM) أجرينا البحث نوعاً وكماً، واستعملنا أربعة وسائل: الاستبيان، الملاحظة، المقابلة، والحوار الجماعي. وقد بَيّنت هذه الدراسة أن النموذج المقترح المتوصل اليه له تأثير إيجابي على الأساتذة والمتعلمين في ممارسة الفعل التربوي بشكل فعال.