

**Democratic and Popular Republic of Algeria
Ministry of Higher Education and Scientific Research**

**Mentouri University of Constantine
Faculty of Letters and Languages
Department of Languages**

Techniques for Measuring Oral Proficiency at University level

**Case Study: First Year L.M.D Students, Department of
English University of Bejaia**

**Thesis Submitted to the Department of Languages in Candidacy for
the Degree of Doctorat d'Etat in Applied Linguistics**

Submitted by:
Mr. DERRADJI Salah

Supervisor:
Prof: ABDERRAHIM Farida

Board of Examiners

Chairman: Prof. MILIANI Mohamed

University of Oran

Supervisor: Prof. ABDERRAHIM

University of Constantine

Farida Member: Prof. SAADI Hacene

University of Constantine

Member: Prof. HAROUNI Zahri

University of Constantine

Member: MC. HOCINE Nacera

University of Annaba

Year 2005

**Democratic and Popular Republic of Algeria
Ministry of Higher Education and Scientific Research**

**Mentouri University of Constantine
Faculty of Letters and Languages
Department of Languages**

Techniques for Measuring Oral Proficiency at University level

**Case Study: First Year L.M.D Students, Department of
English University of Bejaia**

**Thesis Submitted to the Department of Languages in Candidacy
for the Degree of Doctorat d'Etat in Applied Linguistics**

Submitted by:
Mr. DERRADJI Salah

Supervisor:
Prof: ABDERRAHIM Farida

Board of Examiners

Chairman: Prof. MILIANI Mohamed	University of Oran
Supervisor: Prof. ABDERRAHIM Farida	University of Constantine
Member: Prof. SAADI Hacene	University of Constantine
Member: Prof. HAROUNI Zahri	University of Constantine
Member: MC. HOCINE Nacera	University of Annaba

Year 2005



DEDICATION

I dedicate this work:

To my wife Leila

My three sons:

Harris, Sofiane and Rayan

My parents

My parents in law

All my brothers and family

All my friends

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor Prof: Farida ABDERRAHIM for her guidance, her unreserved attention and the time she spent witnessing my efforts throughout this work.

I am also grateful to the teachers of the Department of English, Faculty of Languages and Human Sciences, University of Béjaia whose help and collaboration in the running and answering of the questionnaire is invaluable and to their students who have answered the questionnaire, without whom there could not have been any research.

I would also like to extend and express my thanks to Prof. Mohamed MILIANI from the University of Oran for his helpful discussions and suggestions while doing this work.

My sincere thanks in advance again to Prof.Mohamed MILIANI, Prof. Hacene SAADI, Prof Zahri HAROUNI, Dr .Nacera HOCINE for their acceptance to be members of the jury and for their great efforts and effective cooperation.

ABSTRACT

The present study which is composed of eight chapters, discusses issues related to language testing; covering points which deal with principles of testing, basic considerations and approaches in language testing. Special interest is reserved to examine the development of some adequate oral testing techniques which will develop genuine oral communication in the students of English (first year students, Department of English, Faculty of Languages and Human Sciences, University of Béjaia.)

Investigations and research have shown that in very recent years, there has been a growing interest in the application of testing procedures that are totally different from traditional forms of assessment. More authentic forms of measurement have become increasingly popular in the foreign language classes. These forms of assessment are more student centered in that, in addition to being an assessment tool, they provide the students with instruments to be more involved in their learning "journey", and give them better sense of control and responsibility of their own learning.

It is in this light, that this research work aims to identify, define and propose some of the alternative, authentic forms of oral assessment which can offer to both learner and teacher(tutor) valuable insight into the individual's cognitive processes, rather than simply revealing an ability to memorize and repeat. We believe that the proposed alternatives can better meet the needs of a wider range of students in facilitating learning and readiness for life outside the classroom.

The study discusses the history of portfolio assessment, decisions that need to be made before beginning the portfolio assessment process (for example, what it will look like, who should be involved, what should be assessed, how the assessment will be accomplished), designing a portfolio system (for example., criteria and standards), using portfolio results in planning, and issues related to assessment practices (for example, accountability).

The principle of this new vision is that learners should demonstrate, rather than tell about, what they know and can do. Documenting progress toward higher order goals such as application of skills and synthesis of experience requires obtaining information beyond what can be provided by standardized or norm-based tests. In this "authentic assessment", referred to as "portfolio assessment", information or data is collected from various sources, through multiple methods, and over multiple points in time. Contents of "portfolio assessment" (sometimes called "artifacts" or "evidence") can include drawings, photos, video or audio tapes, writing or other work samples, computer disks, a research work done individually or in pairs in a library or collected from internet or from any other source.

To facilitate the understanding and implementation of this new notion of assessment, both teachers and learners were provided with the needed support and handouts (literature) on how to use portfolios, including their design, management, and interpretation. Suggestions were also offered, aiding students to become better learners. Guidelines were given for planning portfolios, how to use them, selecting of portfolio contents, reporting strategies, and its use in the classroom

The investigation was led through two questionnaires, (one for the students and another for the teachers) in order to build up an expression of

learners' needs and specify the effectiveness of the alternative to be adopted for this population of learners which will promote better learning (by increasing self confidence and motivation in comparison to the traditional way of assessment). After the analysis of the data collected from the two questionnaires and the tutoring sessions, the results revealed that a large majority of the informants (both teachers and students) expressed their readiness and will to adopt the new method, if they are prepared and trained for such challenge. Consequently, a portfolio technique is recommended by the writer as an efficient tool of evaluation, not only to enhance the students' ability to communicate in the target language, but also to gain a progressive autonomy and self reliance in their learning process as well.

LIST OF ABBREVIATIONS AND SYMBOLS

ACT FL	: American Council on the Teaching of Foreign Languages
BBC	: British Broadcasting Corporation
CAT	: Computer Adaptive Testing
CBT	: Computer Based Testing
CLA	: Communicative Language Ability
CLT	: Communicative Language Teaching
CNN	: Cable Network News
CRT	: Criterion Referenced Test
EFL	: English as a Foreign Language
ELT	: English Language Teaching
ESP	: English for Specific Purposes
GE	: General English
IELTS L.	: International English Language Testing System
M.D	: Licence, Master, Doctorat
MCQ N.	: Multiple Choice Questions
R.T	: Norm Referenced Testing
TOEFL	: Teaching of English as a Foreign Language

LIST OF TABLES

Table 1: Mismatch between Learning Activities and Testing Activities	18
Table 2: Superior - Expert User Assessment Scale	119
Table 3: Advanced Level Assessment Scale	120
Table 4: Intermediate Level Assessment Scale	122
Table 5: Novice Level Assessment Scale	125
Table 6: Oral Proficiency in Spoken English.	173
Table 7: Students' Reaction to Testing.	174
Table 8: Students' Nervousness before the Test.	176
Table 9: Students' Nervousness during the Test.	177
Table 10: Students' Opinion about the Teacher being the Sole Responsible 179 for Testing.	
Table 11: Students' Definition of Testing	181
Table 12: Exploiting Testing Sessions to Enhance Students' Oral Performance	182
Table 13: Regularity in Testing	183
Table 14: Reaction to Being Regularly Tested	184
Table 15: Relatedness of Conducted Tests to the Course	185
Table 16: Frequency of Relatedness of the Tests to the Course	186
Table 17: Relatedness of Test Results to Students' True Level	188
Table 18: Frequency of Relatedness of the Results to Students' True Level 189	
Table 19: Decisions Made when the Test Results do not Reflect Students' True Level.	190
Table 20: Possibility of Doing Better in a Test if Conducted by another Teacher	191
Table 21: Appropriateness of the Level of Difficulty of the Test Activities 192	
Table 22: Frequency of Appropriateness of the Level of the Test Activities 193	
Table 23: Adequacy of Discrimination between the Students'	195

Performances at Different Levels of Attainment

Table 24:	Frequency of Adequacy of Discrimination between the Students' Performances at Different Levels of Attainment	196
Table 25:	Adequacy of the Time Allotted	197
Table 26:	Awareness about the Marking Scheme Used by the Teacher	198
Table 27:	Language Aspects included in the Marking Scheme	199
Table 28:	Clarity of Indication of the Marking Scheme in Relation to the Marks to be Awarded for Different Parts of a Question or the Relative Weighting of Criteria that Might be Applicable.	200
Table 29:	Non linguistic Features in the Marking Scheme	201
Table 30:	Frequency of Evaluating Non linguistic Features	202
Table 31:	Techniques Used to Evaluate the Students' Oral Proficiency.	204
Table 32:	Adequacy of the Techniques Used to Evaluate the Students' Oral Proficiency	206
Table 33:	Students' Understanding the Test Directions	209
Table 34:	Students' Familiarity with Self -Assessment	211
Table 35:	Students' Ability of Self —Assessment	212
Table 36:	Opportunities Given to Students to Self —Assess	213
Table 37:	Experience in teaching	220
Table 38:	Experience in Teaching Oral Expression	221
Table 39:	Experience in Teaching Oral Expression in the First Year	222
Table 40:	Teacher the Sole Responsible for Testing?	223
Table 41:	Teachers' definition of Testing	225
Table 42:	Exploiting Testing Sessions to Enhance Students' Oral Performance.	226
Table 43:	Regularity in Testing Students	227
Table 44:	Devising Tests	228
Table 45:	Administering and Conducting Tests	230

Table 46: Teachers' opinion about the Relationship of Conducted Tests to Covered Courses	231
Table 47: Frequency of Relationship of Tests to Covered Topics	232
Table 48: Relationship of Test Results to Students' True Level	234
Table 49: Frequency of Relationship of Test Results to Students True Level	235
Table 50: Frequency of Appropriateness of the Difficulty of the test Activities	236
Table 51: Frequency of Adequacy of Discrimination Between the Performances of Candidates at Different Levels of Attainment	238
Table 52: Relationship between Rewarded Abilities and the Activities Designed to Assess	239
Table 53: Anticipation of Candidates' Responses by the Marking Scheme	240
Table 54: Clarity of Indication of the Marking Scheme in Relation to the Marks to be Awarded for Different Parts of Question or the Relative Weighting of Criteria that Might be Applicable.	
Table 55: Possibility of Alternative Answers by the Marking Scheme	242
Table 56: Subjective Judgement in the Marking Scheme	243
Table 57: Easiness of Interpretation of the Marking Scheme so as to Ensure Reliable Results.	244
Table 58: Order of Language Aspects Tested	246
Table 59: Evaluation of Non linguistic Features	247
Table 60: Frequency of Measuring non linguistic Features	248
Table 61: Techniques Used to Evaluate Students' oral Proficiency.	250
Table 62: Teachers' Opinion about Students' Ability to Evaluate Themselves	251
Table 63: Opportunities Given to Students to Assess their Oral Proficiency	253
Table 64: Possibility of the Administered Tests to Take another Form	254
Table 65: Student Self-Assessment of listening and Speaking	290
Table 66: Proficiency Guidelines: A Self-Assessment of Speaking	291

Table 67: Rubric for measuring students' Oral proficiency	292
Table 68: Rubric for Assessing Portfolio Oral Presentation	293

LIST OF GRAPHS

Graph 1: Oral Proficiency in Spoken English	174
Graph 2: Students' Reaction to Testing	175
Graph 3: Students' Nervousness before the Test	176
Graph 4: Students' Nervousness during the Test	178
Graph 5: Students' Opinion about the Teacher Being the Sole Responsible for Testing	179
Graph 6: Students' Definition of Testing	181
Graph 7: Exploiting Testing Sessions to Enhance Students' Oral Performance	182
Graph 8: Regularity in Testing	183
Graph 9: Reaction to Being Regularly Tested	184
Graph 10: Relatedness of Conducted Tests to the Course	186
Graph 11: Frequency of Relatedness of the Tests to the Course	187
Graph 12: Relatedness of Test Results to Students' True Level	188
Graph 13: Frequency of Relatedness of the Test Results to Students' True Level	189
Graph 14: Decisions Made when the Test Results do not Reflect Students' True Level.	190
Graph 15: Possibility of Doing Better in a Test if Conducted by another Teacher	191
Graph 16: Appropriateness of the Level of Difficulty of the Test Activities	193
Graph 17: Frequency of Appropriateness of the Difficulty of the Test Activities	194
Graph 18: Adequacy of Discrimination between the Students' Performances at Different Levels of Attainment	195
Graph 19: Frequency of Adequacy of Discrimination between the Students' Performances at Different Levels of Attainment	196

LIST OF DIAGRAMS

Diagram 1: Components of Communicative Language Ability in Communicative Language Use.	85
Diagram 2: Components of Language Competence.	86

CONTENTS

INTRODUCTION

I- Statement of the Problem	2
II-Aim of the Study	4
III- Hypothesis	5
IV Means of Research	7
V-Description of the Study	9

CHAPTER ONE: PRINCIPLES OF LANGUAGE TESTING

Introduction	12
I- Testing and Teaching	13
II-Types of Assessment	14
1- Summative Assessment	15
2- Formative Assessment	17
III- Aims of Testing	22
1-The role of Testing	22
2-Learners' Expectancies	25
3-Teachers' expectancies	26
4-Evaluation of the programme	28
5-Institutional requirements	29
IV- Characteristics of Tests	31
1- Test Functions	31
2-Objective and Subjective Testing	32
Conclusion	36

CHAPTER TWO: BASIC CONSIDERATIONS IN TEST DESIGN

Introduction	38
I- Developing a Test	38
1-Stages of Development of a Test	38
2-Test Items	40
3-Variables Affecting Test Performance	42
II - Criteria of Test Design	44
1- Validity	45
1.1- Content Validity	46
1.2-Criterion Validity	51
1.3-Construct Validity	56
1.4-Wash back Validity	61
2-Reliability	64
-Factors which Affect Reliability	65
3-Practicality	69
Conclusion	70

CHAPTER THREE: APPROACHES TO LANGUAGE

TESTING

Introduction	72
I -The Structuralist Approach	73
II- The Psycho- Sociolinguistic Approach	77
III-The Communicative Approach	82
1- Communicative Language Ability	82
2- Characteristics of a Test	87

Conclusion	-	97
------------	---	----

CHAPTER FOUR: TESTING ORAL PROFICIENCY

Introduction		99
I- The Oral Proficiency		100
II- Oral Proficiency Testing		103
1-Approaches to Testing the oral proficiency		103
2- Specificity of Oral Proficiency Testing		106
3-Criteria of Testing Oral Proficiency		107
4- Improving the Reliability of Oral Examinations		111
5- The Assessment Scales		118
III- Procedures in Testing Oral Proficiency in the Department of English		127
Conclusion		129

CHAPTER FIVE: ORAL PROFICIENCY TESTS

Introduction		131
I- Traditional Assessment Techniques		131
1- Multiple Choice Questions (MCQ)		131
2- The Oral Interview		145
3- Group Work		138
4- Problem-Solving		140
5- Role Play		142
6- Discussion		144
II-Self -Assessment		146
1-Definition		146
2-Research in Self- Assessment		147
3-Aspects of Self -Assessment		150
3.1-Responsibility		151

3.2-Learner Training	152
3.3-Innovation in Self Assessment	153
4-Reservations as to the use of Self Assessment: The Problem of Bias	154
III-Portfolio Assessment	156
1-What is a Portfolio	156
2-Why Use Portfolio Assessment	157
3-Stages in Implementing Portfolio Assessment	159
4-Advantages and Disadvantages of Using Portfolio Assessment	164
Conclusion	167

CHAPTER SIX: NEEDS ANALYSIS

Introduction	169
I- Students' Questionnaire	170
1-Aim of the Questionnaire	170
2-Description of the Questionnaire	170
3-Analysis of the Results	173
II-Teachers' Questionnaire	216
1-Aim of the Questionnaire	216
2-Description of the Questionnaire	216
3-Analysis of the Results	220
Conclusion	256

CHAPTER SEVEN: PORTFOLIO ASSESSMENT

Introduction	258
I- Developing a Portfolio	258
1- Directions to the Students	258
2- Students' Portfolios	262
II- Scoring the Portfolios	288
1-Criteria of Portfolio Assessment	288
2-The Rating Scale	294

III- Evaluation of the Experiment	297
Conclusion	299
<u>CHAPTER EIGHT: RECOMMENDATIONS</u>	
Introduction	301
I- Designing a Portfolio	301
II- Answers to Teachers' Concerns about Portfolio Assessment	304
III- Answers to Students' Concerns about the Oral Tests	309
IV- Advice and Action Plan for Students	314
V- Advice to Teachers	316
Conclusion	318
GENERAL CONCLUSION	319
APPENDICES	323
APPENDIX I: The LMD System	324
APPENDIX II: Students' Questionnaire	334
APPENDIX III: Teachers' Questionnaire	348
GLOSSARY OF IMPORTANT TESTING TERMS	362
BIBLIOGRAPHY	371

INTRODUCTION

I - Statement of the Problem	2
II - Aim of the Study	4
III- Hypothesis	5
IV-Means of Research	7
V-Description of the Study	9

INTRODUCTION

I- Statement of the Problem

The assessment of language learning results, especially in a formal testing situation, has become more difficult than it used to be. There are several reasons for this. The most important one is the shift of emphasis from reading and writing skills, or from formal correctness in the use of the language, to aural/oral skills and to what has become commonly termed «communicative competence ». In the traditional school setting, for instance, to be able to translate prose texts into the foreign language, to be able to produce a neatly organised essay on a specified topic, are manifestations of language use which are comparatively simple to assess, by means of ordinary paper-and-pencil tests or conventional written exams. In this kind of tests, testers can administer the test to many people at once, do the evaluation (the marking) all by themselves, without using a complicated technical apparatus, and do it at a time which suits them.

Measuring oral proficiency (« communicative competence »), i.e., the skill which most learners today regard as their first priority is much more problematic, undertaking, especially when the teacher is faced with large sections. It requires individual treatment of the learners (although speaking tests are sometimes administered as group tests in language laboratories), and a fairly elaborate testing arrangement, typically with a trained tester conducting a structured interview. This means that one has to solve a number of practical problems (for example, the number of markers, the marking

scheme) and the need to have more time at one's disposal. There are also some very difficult criterion problems to tackle: in which case did communication actually take place and to what extent should one take into account non-linguistic features, i.e., gestures, facial expressions, and other non-verbal signals?

When assessing Oral Proficiency, one has always to find answers to questions related to validity:

How does one measure linguistic aspects such as fluency, stress, intonation, correctness and appropriateness of response. (Which, needless to say, are very important from the communication point of view)? How should these components of the spoken language be weighted against more easily evaluated variables such as syntactic accuracy, vocabulary command, pronunciation correctness?

Further, there is the problem of how to make sure that different testers (for example, interviewers) apply the same standards when scoring. There must be some guarantee that a given candidate is judged in roughly the same way whoever conducts the test. This relates to the reliability problem, which can only be solved through extensive experimentation and careful statistical analyses.

In sum, we can say that nowadays the importance of aural/oral skills and «communicative competence » is emphasised in language learning and testing. In practice, this means that it has become more difficult for the language teacher and tester to assess the results of learning (mainly because oral tests are more time-consuming to administer and more difficult to evaluate than conventional tests measuring other skills).

II-Aim of the Study

In this study, we aim to demonstrate that some of the alternative forms of oral assessment can offer both learner and teacher (tutor) valuable insight into the individual's cognitive processes, rather than simply revealing an ability to memorize and repeat. If this focus on the process of learning begins early in the programme of study, it means that the benefits of understanding one's own learning agenda are felt from the outset. It also allows time for the student to further develop as an effective learner, over time, gaining proficiency in the range of learning skills most appropriate to his / her own context.

Our interest centres around testing oral proficiency. Testing this ability requires specific techniques: Multiple choice questions, group work, role play, problem solving, discussion, self assessment and portfolio. In our magister study, we came out with the conclusion that learners will be able to speak comprehensibly if some of the stated techniques are used to enhance the skill and engender spontaneous flow of language. Teachers can find and adopt more and more interesting, genuine and adequate methods of developing the spoken language they can try to apply these in the classroom, advocating authenticity of materials, relevance of situations in which speaking is an appropriate activity, reflecting cultural sensitivity. They can make the learning environment as conducive to expression and language acquisition as possible

In this study, we are interested in investigating the use of the proposed techniques and their appropriateness, with the ultimate aim to propose to first year university students at the Faculty of Languages and Human Sciences, University of Béjaia, a battery of tests which could be used in the context of the listening and speaking comprehension unit. The students involved in this

study, are enrolled in the new system of higher education, i.e. Licence, Master, Doctorat. (LMD). This new proposed system of education is based on international norms; it is a response to the dysfunctions which appeared in the previous system. This new vision aims to ensure a high quality education. (See Appendix 1)

III-Hypothesis

We believe that our students need to be trained so as to make of them independent, conscious and confident learners. **In** this sense, we hypothesize that we can help them increase their motivation and personal involvement in the learning process if we introduce the strategy of assessing one's abilities through the Portfolio Assessment Project at different points in the language learning process. The teacher (the tutor) has also a very determinant role in incorporating these strategies as cornerstones in the student's learning process and environment.

Students need to self-assess in order to monitor their learning progress. For example, through portfolio and self-assessment, students will learn that their progress and grade do not depend solely on the teacher, but in large part on themselves. Regularly used, portfolio and self-evaluations can help students track areas of strength and weakness and address problems before final exams are issued. Self-evaluation also allows teachers to see how students view their progress, leading to individualized instruction and to initiation to autonomous learning.

It is true that many students may not at first feel comfortable with self-evaluation, thinking that it is the job of the teacher to evaluate. However, it is our strong belief that students cannot grow as active learners if they are not able to assess their own abilities and progress and set a path for future

learning. Students need to be taught the strategy of self-assessment explicitly and to be supported in their use of the strategy. The greatest value of self-assessment and portfolios is that, in building them, students become active participants in the learning process and its assessment.

We also believe that promoting this new vision to assessment will hopefully engender a dynamic pedagogy where the students are actors of their training and the educational team is a support, a guide and adviser which come with them all along their academic training. When students are asked to rate their current abilities, the academic team (tutors) will be prepared to suggest and provide them with the tools and criteria with which they will assess themselves.

IV-Means of Research:

In order to propose an effective battery of tests for a population of first year students of English, we started by assessing the effectiveness of the techniques which were being used to measure the learners' oral proficiency and suggest alternatives of measurement. We made recourse to two means of research commonly used in second language studies.

- Students' and teachers' Questionnaires to examine the existing system of evaluation and find out whether the population under study accepts alternatives and the new vision, in the context of oral proficiency assessment.
- Experimentation done on a group of first year students, who are encouraged and guided to make their own portfolios.

The oral proficiency of the population under study was evaluated through "a Portfolio". We introduced the idea of "a Portfolio" to the learners which is the selection of the student's work that is a representation of his/her development as a speaker, language user, and thinker over time. We also explained that there were two kinds of portfolios: comprehensive and selective. A comprehensive portfolio contains all student's work, including notes, diagrams, peer responses, final revisions, and teacher comments. A selective portfolio does not contain every artifact that he/she has produced in the class over the course of the semester, but rather a selection of work that he/she feels best represents his/her progress in this course as a user of English. In this study, we opted for a selective "Portfolio".

A few weeks before conducting the test, we made it clear to the students involved in the study that what they need to do is to include artifacts in their portfolios that evidence their development with respect to the study of English and their reflections on that development. We also explained to them

that they should not necessarily place their best work in their portfolios, but the work that most clearly represents their development, and that if they simply put copies of assignments into their portfolios, they will not receive a favourable evaluation. They were encouraged to reflect on their work and comment on why they think particular artifacts represent their development and progress. This means that for each sample of their work that they place in their portfolio, they must also place an accompanying piece that explains in detail how the first piece represents their development.

We also clarified to the students that comparison and reflection are crucial elements of "a portfolio". For example, they may choose to place the first and second drafts of their first recording of a passage or a discussion in their portfolio. Alongside these two artifacts, they must place their reflections on their development from the first version to the final version. They might include concrete examples of how they began to introduce themselves during a summer course in an English speaking country, using different expressions for introducing themselves. They might also discuss content realizations that they had in the course. We proposed to the students, examples of the types of data they might place in a portfolio, such as: a recording of their speaking English, reciting a poem, the results of an office (tutoring) hour which they spent with their professor, the notes they prepared for class discussion as they were listening to evening news on BBC or CNN..... Explaining that the principle of a portfolio is that they can place anything of interest in there.

In such measurement, students are not graded against a norm or standard. If we accept that we are all individuals, we all have different abilities, capabilities, interests, strengths and weaknesses. Therefore, it is not fair to subject the learners to a one-size-fits-all assessment such as a grammar quiz or a multiple choice exam. Instead, in this type of assessment learners

have the opportunity to capitalize on their strengths, interests, and abilities in the form of portfolio assessment.

V-Description of the Study

The study is divided into eight chapters. The first chapter presents an account of the principles of language testing. It deals with issues related to types and aims of testing. It also discusses the characteristics of different tests.

The second chapter reviews the basic considerations in test design. It covers the concepts of validity, reliability and practicality. The status of the various types of validity, and how the concept of validity relates to those of practicality have been examined from a deeper, theoretical perspective.

Chapter three is an attempt to illustrating the different approaches to language testing and their limitations in terms of the criteria of validity, reliability and practicality. These approaches to testing are described in terms of the particular focus they represent.

The fourth chapter provides the methodology related to approaches and aims of testing the oral proficiency. We have highlighted some criteria to consider while testing the oral ability, taking into consideration elements such as reliability and validity. Some assessment scales of this proficiency are proposed as well.

In the fifth chapter, different techniques for testing oral proficiency have been appraised, mentioning the element of innovation. Although their coverage is wide and interrelated, only their applicability in the language

classroom has been considered. Some of the techniques proposed promote a new vision of measurement based on self assessment.

Chapter six, which is the first practical part of this study, is devoted to the analysis and interpretation of the data collected from the students' and teachers' questionnaires.

Chapter seven is based on the description, analysis and assessment of the "portfolios" produced by the students. It illustrates the principles of this alternative assessment and the different steps to follow.

Finally, in chapter eight, we have provided the reader with some suggestions and recommendations to the implementation of self assessment and portfolio.

It was also felt necessary to provide the reader with the most important testing terms and definitions that should prove useful in helping to understand the information provided in this research work.

CHAPTER ONE

PRINCIPLES OF LANGUAGE TESTING

Introduction	12
I- Testing and Teaching	13
II-Types of Assessment	14
1-Summative assessment	15
1-Formative assessment	17
III- Aims of Testing	22
1-The role of Testing	22
2-Learners' Expectancies	25
3-Teachers' expectancies	26
4-Evaluation of the programme	28
5-Institutional requirements	29
V- Characteristics of Tests	31
1- Test Functions	31
2-Objective and Subjective Testing	32
Conclusion	36

PRINCIPLES OF LANGUAGE TESTING

Introduction

The amount of work which has been done in the testing field is quite substantial. Most of the research, discussion, thinking and writing, however, has focused on testing methods, i.e., the how of testing. On the other hand, the why of testing —perhaps not surprisingly- is seldom questioned.

We can say that evaluation, in a broad sense, is part of communication. When two or more people interact, their language production is a result of choices from a language repertoire, according to the different factors at play in a given speech event in Hyme's terms, i.e., the setting, the topic, the purpose of interaction, the roles of addressor and addressee, the channel and the message form. On the basis of these assumptions, each participant in the interaction will probably judge his/her own decisions regarding how to proceed in the interaction (Breen and Candlin: 1980). Other sorts of behavior —not necessarily linguistic- are also evaluated on the basis of the same assumptions, for example, gestures, attitudes, way of dressing. Clearly enough, such an evaluation is usually not overtly expressed. Moreover, it is true that different participants in an interaction might be perceived as having more or less status, power or authority.

This is the sort of evaluation which happens in "normal" conversation. When we look at the context of the foreign language classroom, however, as in most school situations, the concept of evaluation is narrowed down to become a synonym of formal tests. This idea is so established as part of school life that tests are an expected component of a school curriculum.

I-Testing and Teaching

We can notice that so many examinations in the past have led to a separation of testing from teaching. However, we believe that both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other. Teaching has always been a process of helping others to discover "new" ideas and "new" ways of organizing what they learned. Whether this process took place through systematic teaching and testing, or through a discovery approach, testing was, and remains, an integral part of teaching.

According to Mac Donald, (1973:1-2) "Testing is the process of conceiving, obtaining and communicating information for the guidance of educational decision making, with regard to a specified programme" As for Rea-Dickins and Germain, (1992:16) "Testing is the principled and systematic collection of information for purposes of decision making" Earlier definitions of evaluation tended to focus on judgemental aspects and reinforce, the potential of evaluation in a supportive role vis a vis institutional, curriculum development.

Tests may be constructed primarily as devices to reinforce learning and to motivate the student, or primarily as a means of assessing the student's performance in the language. In the former case, the test is geared to the teaching that has taken place, whereas in the latter case the teaching is often geared largely to the test. Standardised tests and public examinations, in fact, generally exert such a considerable influence on the average teacher that they are often instrumental in determining the kind of teaching that takes place before the test.

It can be argued with some justification that language examinations in the past have exerted a harmful influence on the language teacher and have

considerably inhibited language learning by encouraging teachers to teach *about* the language. As a consequence, relatively few teachers sought to provide the maximum opportunity for their students to practise the language itself. Many external examining bodies today seek to measure the candidate's actual performance in the language, and in this way, sometimes, exert a beneficial influence on syllabuses and teaching strategies. Yet, however much concerned a public examining body may be about the effects on teaching (i.e. the backwash effects which will be discussed in detail in chapter two) of its particular examination, the main purpose of that examination is to measure the candidate's ability to use the target language. The examination is thus primarily an instrument for measuring and evaluating performance.

In this chapter, we will look at some of the possible reasons why formal tests have such an important role in most school systems. We will also discuss the implications of an approach to evaluation which disempowers both learners and teachers and does not encourage learner responsibility and autonomy.

II-Types of Assessment

The expression "formal test" has been used to refer to the kind of test which is set and recognized by an institution of examining board for the sake of placement, promotion or the issuing of a certificate. By contrast, informal tests would be the small-scale, teacher-prepared tests which are used as spot-checks during a course of instruction. The latter may not necessarily be used for the sake of attributing a set of marks to each learner but would mainly serve as feedback for teachers and learners. Another way of putting these different sorts of tests into perspective, especially in relation to the frequency with which they take place is in terms of summative and formative assessment. Bloom (1971:62) defines summative assessment as "a judgment about learners with regard to

the effectiveness of learning and instruction after it has taken place").and Formative assessment, as "the use of systematic evaluation during the process of learning for the purpose of improving that process". From this definition, we can see how these types of assessment are usually viewed and implemented.

1-Summative Assessment.

Summative assessment usually takes the form of a "final" test, i.e., one which is taken by students at the end of a certain course period—a month, a course book, a semester, an academic year, a whole course. It aims at evaluating learners' performance, attributing them a set of scores and a pass or fail mark. This test is written by the classroom teacher or most probably by an examining board, either external to or from within the institution and it is normally considered to be an objective, valid measure of performance. Summative assessment usually gives learners little or no feedback in terms of particularly good areas or weak areas of performance where improvement is required.

The main point to be mentioned in favor of summative assessment is that, it is a way of checking individual learners' performance in relation to course objectives. We can say, however, that this presupposes tests which attend to the demands of construct and content validation as well as those of reliability. We must have very well defined objectives for a course as a whole and for the different segments of this course, for example, semester, months. If all these conditions are met, summative assessment has a clear role to play in an institutional context. This role is basically that of fulfilling the requirements of the different (sometimes opposing) forces of the different parties involved in a testing operation, mainly those associated with an institutional perspective.

The problems or disadvantages of summative assessment are basically a matter of emphasis of both the frequency with which summative assessment occurs and the weight it has in relation to formative assessment.

Rea (1985:30) defines evaluation as testing activities which satisfy requirements external to the teaching situation, i.e., those identified by an institution or examining board. "Assessment and appraisal activities are, on the other hand, internal to the English Language Teaching (E.L.T) programme. The former generally implies assigning a grade to a piece of work whilst the latter does not". From Rea's point of view, we can imply that the overall grade is a sum of the different marks obtained by a candidate in different instances of evaluation. In other words, assessment and appraisal seem to have the function of providing learners with feedback on their progress, but not marks upon which their promotion or issuing of degrees may be decided. It is also interesting to notice that any form of evaluation which is carried out by students —i.e., self-appraisal or peer -observation- does not involve the assignment of grades to students' work.

We would find it difficult to convey to students (and teachers) ideas such as students' responsibility for their own learning and a focus on the process of learning.(these are indeed the recommendations of the LMD system) If only the grades attributes to formal tests described as "external" to the teaching process (in the sense that they are probably standardized and imposed from authorities which are above the teacher) are taken into consideration for the issuing of the overall grade, teachers' and students' assessment is automatically discredited.

Moreover, if these tests are implemented at very short intervals, one has to assume that learning objectives can be divided into equally short, self-contained units. Learning would consist of mastery of separate blocks which should be

mastered over a certain period of time and immediately put into practice or `display" through a test.

2-Formative Assessment

Formative assessment is basically aimed at improving the process of learning. Improvement will be sought through actions in relation to feedback given to teachers and learners during the course of instruction. This feedback can be collected through, not only typical test-like devices, but also through teachers' observations, questionnaires, peer and self-evaluation. The marks attributed to such exercises are not necessarily added to students' overall scores. According to Rea(1985:29)

. .testing strategies for the purposes of formative assessment will include self-appraisal, peer group observation, and record and diary keeping. It will be noted that student participation is largely confined to appraisal activities whereas assessment in the form of test will be largely the domain of the teacher.

Formative assessment aims at identifying areas which might benefit from intervention and therefore lead to improvement. The different instruments used for assessment for example, homework, performance in classroom activities, responses to questionnaires- will try to provide learners and teachers with insights into individual learners' development in relation to the programme. Adjustments will then be made, either in the programme or by individual students, in case problems are spotted and not a pass or fail, after which there is nothing else to do except for going on to the next level in the course or repeating the same one.

<u>Proficiency-oriented learning activities</u>	<u>Current achievement testing activities</u>
- Large variety of tasks and contexts	- little variety in test methods and formats
- Integration of language skill areas	- Focus largely on separate receptive skills
- Synthesis of knowledge and skills in extended and global tasks	- diagnostic discrete-point tasks and scoring procedures
- Focus on message, function and form	- focus form before function and message
- Group collaboration and individual work	- individual work only (group effort=cheating)
- encourage resourcefulness in resolving authentic problems in language use (e.g. cognitive and international ones)	- encourage accuracy in resolving contrived problems that arise at the linguistic level but rarely at cognitive, affective and other levels
-Focus on the learner as a responsible agent of his or her own learning	- treat the examinee as one who simply takes tests but does not contribute to their design and use
- exploit assessment as an inherent and natural activity in any authentic learning and use of language.	- treat assessment as obligatory and scheduled activity imposed on learners after authentic learning and use of language take place.

If formative assessment is to account for and promote individualization, one should assume that self-access work done outside classroom hours will be encouraged as is the case for LMD (Licence, Master and Doctorate) students in the Department of English, Faculty of Languages and Human Sciences University of Béjaia. This brings us to the idea of learning skills or strategies, and a greater awareness of learners in relation to their own development. This move certainly presupposes a good amount of guidance (tutoring) on the part of the teacher and probably a more collaborative classroom atmosphere, involving not only teacher-student interaction but also peer-criticism as well. It follows then; it seems to us, that teachers and learners have to be empowered with the official recognition of their assessment of learners' work. In order to do so, it is crucial that formative assessment be given not only credit but substantial weight within the overall assessment system.

When discussing the uses of formative assessment, one cannot forget the conflict between individualization and standardization, and the issue of responsibility. The first problem, although being by no means an easy one to solve, is perhaps at least more easily identifiable. On the other hand, if this institution has a broad educational perspective rather than a narrow-minded training programme of language teaching, standardization becomes problematic. We come back then to the issue of balance between formative and summative assessment, and to a programme of instruction which is clear in terms of objectives but which offers a good deal of flexibility in terms of the means to achieve them.

The second problem, i.e., that of the responsibility for learners' development and assessment, is both tricky and less easily identified. In the first place, being responsible or being made responsible for oneself or others is a task which some people may not be willing to accept; this what the Department of

English is experimenting. It usually implies doing hard work and having to respond for one's actions. This is quite different from sitting in a classroom and being told what to do, or sitting for a test which somebody else has prepared and whom one can blame in the case of failure. In assessment based on class or homework, the problem of teachers' responsibility must also be handled with care. It is true that one of the advantages of formative assessment is to promote more involvement of teachers and learners in the learning process. But, if the teacher's assessment has considerable weight on the students' overall grades, two problems may arise. On the one hand, the position of the teacher as a figure of authority in relation to students is reinforced, therefore affecting the "team spirit". On the other hand, teachers may perceive the responsibility of attributing marks which might determine whether a student should pass or fail as an extra burden and one which is too heavy. They might feel that this is a task to be undertaken by the institution, as a sort of anonymous figure. The reasons for this resistance may be varied, for example:

- If there are too many students in a group, having a fairly accurate judgment of each individual student's production in class becomes problematic;
- If teachers do not have clear guidelines in relation to how to conduct continuous assessment, they may feel at a loss, especially if they are not very experienced which is indeed the case in the Department of English adopting the LMD system;
- In the case of teachers of young learners in particular, an emotional attachment usually emerges out of the work developed along an academic year. If getting a low overall mark at the end of the semester means that the student will not "pass", and if there is a high social pressure on students in relation to getting pass or fail grades, the teacher will,

understandably, like to have his/her opinion backed up by somebody else. (this often happens at the end or the beginning of the academic year).

The conflict between individualization and standardization and the teachers' responsibility seem to be hard to resolve. At the bottom of this conflict lies the question of responsibility for students' progress and of the power and authority to determine whether students have been successful or not. This last point has also to do with the aspect of credibility. For the general public, formal tests of any sort usually have the credibility which is conferred to them by some sort of authority. Other forms of assessment, however, especially if they include learners' self assessment may encounter a good deal of resistance. They will possibly be considered "too subjective" or "biased", which is not totally a wrong assumption, if continuous assessment is not based on clear, well defined principles.

Educational institutions are trying to promote a greater deal of learner awareness and involvement in the process of learning, with more emphasis on teachers' continuous assessment of learners' performance and consequently, less emphasis on formal end-of-term test. In most settings, continuous assessment means having a formal test after different units of a course book or at the end of each semester. The idea is one of testing achievement at short intervals and adding all these individual marks at the end of a course of an academic year.

The introduction of such a formative assessment followed an approach similar to that proposed by Trump (1967, quoted in White 1988.). After consultation with teachers and planning of the whole project, the implementation of the new format was —and has been up to now- accompanied by careful guidelines and information to students, parents and teachers, as well as monitoring. (the new system of the Licence LMD).The assessment system

being currently used comprises continuous assessment and end-of-term tests. One of the aims of continuous (or formative) assessment in this context is promoting more involvement and participation on the part of learners. The idea is that they are evaluated on the basis of their development and performance throughout the term and not according to the results of a single test. Continuous assessment should also provide learners with consistent and objective feedback on both classroom activities and homework.

Rather than being mutually exclusive, as the title of this section might suggest, it seems to us that summative and formative assessment should be seen as complementary sorts of measure. Balancing the two, nevertheless, would have to be done if one is to attend to the requirements, but having a look at the advantages and disadvantages of both summative and formative assessment may be helpful.

III- Aims of Testing

1-The Role of Testing

Brown, et al (1994: 5) reminds us that it is important as providers of education to step back and consider why we assess. There are times when we become so immersed in the job that we lose sight of the exact purpose of a particular element of assessment. There is then the possibility that we are not achieving that purpose, or that we overlook another form of assessment which might be more appropriate. This last comment is particularly relevant to the assessment of oral competence where there are a range of assessment strategies to choose from.

Brown, et al stated some good reasons why assessment is useful:

- Motivation.
- Creating learning activities.
- Feedback to the student, identifying the strengths and weaknesses.
- Feedback to the staff on how well the message is getting across.
- To judge performance (grade/degree classification).
- Quality assurance-internal and external to the institution.

Tests can provide valuable information about an individual's competence, knowledge, skills, or behaviour. They are based on samples of the many questions which could be asked of the students' behaviour, where we value correct answers and penalize error. What is then the role of tests in program implementation and evaluation more specifically? Tests can be used to:

- conduct needs assessments which determine whether a special program is needed, and if so, what kind;
- plan a specific content of a program;
- select those students who need a special program;
- determine when a student may no longer need a special program;
- determine whether students in the program are progressing as intended;
- determine which parts of a program may need to be revised; and
- demonstrate program effectiveness.

To understand the role of tests, we must also understand much about types of tests and types of test scores, as well as about some of the technical aspects of testing.

The development of modern linguistic theory has helped to make language teachers and testers aware of the importance of analysing the language being tested. Modern descriptive grammars are replacing the older, Latin-based prescriptive grammars: linguists are examining the whole complex system of language skills and patterns of linguistic behaviour. Indeed, language skills are so complex and so closely related to many other non-linguistic skills (gestures, eye-movements, etc.) that it may often seem impossible to separate them for the purpose of any kind of assessment. A person always speaks and communicates in a particular situation at a particular time. Without this kind of context, language may lose much of its meaning. Before a test is constructed, it is important to question the standards which are being set. What standards should be demanded of learners of a foreign language? Should our language learners, for example, be expected to communicate with the same ease and fluency as native speaker? Are certain habits of second language learners regarded as mistakes when these same habits would not constitute mistakes when belonging to native speakers? What, indeed, is "correct" English? Examinations in the written language have in the past set artificial standards even for native speakers and have often demanded skills similar to those acquired by the great English essayists and critics. In imitating first language examinations of written English, however, second language examinations have proved far more unrealistic in their own words or to write original essays in language beyond their capacity.

In the context of a large language educational institution, (case of the Department of English) the role of formal tests is both to provide learners with some sort of certification of proficiency in the foreign language and to provide feedback. However, feedback can only be understood in terms of who gives feedback to whom and of what feedback is used for. One way of identifying and grouping the different parties involved in testing and in feedback

consolidate the pattern taught. In this way, a good test can be used as a valuable teaching device.

In relation to progress tests, according to Skehan (1989) it is easier to identify different degrees of anxiety, which some students generally show before or during the test. Common sense and experience, in this respect, would probably suggest that while some anxiety might be a sign of awareness and commitment to a testing situation, making one perform at his/her best, too much nervousness might hinder performance and should therefore be avoided (Skehan op. Cit :115).

It is our impression that if too much emphasis is placed upon summative assessment, with formal tests being implemented at very frequent intervals; we run the risk of, in the first place, decreasing the level of learner involvement in the process of instruction. Secondly, if tests are not perceived to be a true and clear reflection of the teaching/learning process, they are likely to cause anxiety. Another problem of an excessive emphasis on formal tests refers to the nature of the feedback which is given to the learner. If all that students get is a total score and perhaps brief comments from the teacher and if the same procedure is adopted throughout the different periods of the course, it is fair to think that formal achievement tests usually give little or no indication to learners as to their weaknesses and strengths and consequently will not be good sources of feedback for the learner.

3. Teachers' expectancies

As far as teachers are concerned, tests may be considered as instruments to evaluate not only learners' performance but, in an indirect way, to measure

successful teaching, or teachers' performance. This is the case when tests are not designed by the classroom teacher him/herself.

In terms of feedback, tests can help the teacher in different ways, one of them being a way of making decisions in relation to the promotion of students. Formal tests often provide teachers with a way of validating their own assessment of learners' performance in a course. A correlation between the two forms of assessment becomes increasingly important depending on the decisions which are taken on the basis of test scores, like the issuing of diplomas and certificates at the end of a course. It becomes clear then that teachers must feel confident about their own assessment of learners, based on the latter's production during the whole semester.

This implies that if continuous assessment of learners' performance is part of the overall grade that students get on a course, teachers should have very clear guidelines of how to assess students' performance in order to have a good measure against which formal tests could be validated.

Although most teachers wish to evaluate individual performance, the aim of the classroom test is concerned with evaluation for the purpose of enabling teachers to increase their own effectiveness by making adjustments in their teaching to enable certain groups of students or individuals in the class to benefit more. Too many teachers gear their teaching towards an ill-defined 'average' group without taking into account the abilities of those students in the class who are at either end of the scale, facing serious difficulties. A good classroom test will also help to locate the precise areas of difficulty encountered by the class or by the individual student. Just as it is necessary for the doctor first to diagnose his/her patient's illness, so it is equally necessary for the teacher to diagnose and find out his/her student's weaknesses and difficulties related to the process of

learning. Unless the teacher is able to identify and analyse the errors a student makes in handling the target languages, he will be in no position to render any assistance or guidance at all through appropriate anticipation, remedial work and additional practice.

The test should also enable the teacher to ascertain which parts of the language programme have been found difficult by the learners. In this way, the teacher can evaluate the effectiveness of the syllabus as well as the methods and materials he/she is using. The test results may indicate, for example, certain areas of the language syllabus which have not taken sufficient account of the learner's difficulties or which, for some reason, have been glossed over.

4-Evaluation of the Programme

Another use of the feedback provided by formal tests concerns the evaluation of the programme; especially if they are not prepared by the classroom teacher, formal tests can serve the purpose of throwing light on a programme of instruction. In the case of the Department of English, Faculty of Languages and Human Sciences, University of Bejaia for example, the formal tests which are implemented at the end of each semester are designed either by a group of coordinators in charge of planning all aspects of the different courses, choosing and adapting materials, or designed individually by teachers. Teachers participate in the process of test design at different stages, from analyzing first drafts of test to providing feedback after their implementation. Changes may then take place not only in relation to the tests themselves but also in relation to course programme and materials used.

Evaluation and change in the programme can also be done on a micro-scale, by the teacher him/herself. Since tests cannot possibly cover a

whole teaching syllabus in detail, teachers might use test questions as a guideline to points in the syllabus which should be emphasized. This may be a positive influence of tests, if their format is conducive with the methodology favored by the institution, and if they truly measure achievement. On the other hand, if the teaching syllabus is broader than the testing syllabus, an emphasis on some parts of that syllabus to the detriment of the others is not desirable. This brings us to two points: firstly, the extent to which a test can satisfactorily sample a given syllabus, and the extent to which test items lend themselves to some sort of "specific" training rather than requiring learners to apply general abilities in the foreign language. Secondly, the wash back effect (to be explained in the chapter two) of tests into teaching has to presuppose the readiness on the part of teachers to perceive and carry out processes of innovation.

5-Institutional Requirements

At the institutional level, formal tests serve a multitude of purposes, such as making a selection to offer a job, a promotion or a distinction. This makes one think that tests may be used more as a response to institutional constraints than as an aspect which is vital to the teaching process.

Because they are regarded as being objective, valid measures of performance and because they are usually issued by some sort of authority in a given field —be it a teacher, an institution or an examining board- tests are highly valued, and their results are seldom questioned by the general public. In fact, formal tests have such firm roots within most educational contexts that any form of innovation in the field if it involves suppressing or even reducing the weight of formal tests within an assessment system is regarded as suspicious.

Learners, their parents or sponsors and society in general put a good amount of pressure on an institution in relation to having test results which can be clearly and easily understood as a reflection of progress or achievement in a course. These results should also be documented, if possible, by some concrete evidence like a paper the student has taken. It could be argued that any institution should not base decisions concerning educational policies entirely on the expectations of their public. However, these have also to be taken into account, if the school is to satisfy the market's needs.

Another function of tests, seen from an institutional point of view, is that of their washback effect into teaching. Even when planning and test design are centralized within an institution and imposed upon teachers, tests will inevitably guide teaching (Swain, 1985:43). This is not in itself entirely harmful and can effectively lead to change: more often than not, institutions make use of tests as a means of promoting innovation or facilitating the implementation of new ideas in relation to pedagogic practices. In some cases, institutions make use of tests as sources of feedback about the teaching process as a whole. It is also acknowledged that test implementation and test results can provide course planners with valuable insights into the suitability of materials and methods, as well as help them to evaluate course objectives. In this respect, the more centralized the decisions about the teaching process are, i.e., the more it approaches a top- down model of decision-making, the more likely it will be for tests to be seen as instruments of control over the teaching process, and ultimately, of teachers. To understand the role of tests, we need to understand much about types of tests scores, as well as about some of the technical aspects of testing.

IV-Characteristics of Tests

1- Test Functions

A look at different kinds of tests, defined in relation to their function, may prove useful for the discussion of what we should test, or rather, what we should be looking for when analyzing test results. The classification provided by Davies (1977) identifies five types of language tests: achievement, proficiency, aptitude, diagnostic and progress.

Achievement (or attainment) **tests** are defined against a set of identified and pre-determined objectives of a course of study. They reflect the propositions of the curriculum in a given school or educational system. The main aim of achievement tests is to measure the extent to which each learner has attained some specified objectives, or in other words, whether an individual "has learnt" what s/he was taught. They are, in this sense, very much linked to what happened before the test, during the course of instruction.

Proficiency tests are defined in relation to the knowledge of a foreign language judged necessary for the performance of certain tasks, or judged to match some sort of "native-like" abilities. Proficiency tests are designed irrespective of the teaching/learning experience candidates have been exposed to and are geared towards what candidates are supposed to be able to do after the test.

Aptitude tests are mainly pre-course instruments designed to establish how successful an individual might be in learning a language.

Diagnostic tests are small-scale, usually teacher-prepared instruments supposed to serve as feedback for teachers (and learners). They should spot

problematic areas which might benefit from remedial work. They can also be used to:

- identify students' strengths and weaknesses in specific content areas,
- determine how best to help the student overcome any particular limitations, and
- provide norms against which a student can be compared in order to determine which skills should be taught (as with standardized achievement tests, placement/diagnostic tests).

Progress tests are intended to show the extent to which an individual has been keeping up with the work done on a certain course. They certainly have a diagnostic function as well, but will probably be used as the criterion for promotion along different levels of a course.

Although test design is not necessarily a direct reflection of the purpose which a test is supposed to serve, it is important to have the different functions in mind, given the prominence of formal tests in most educational systems. In terms of the issues discussed in test function: it seems that a combination of progress and achievement tests might be desirable. The purpose of such tests is mainly that of providing a basis for the assessment of students' progress and for decisions concerning their promotion to higher levels. Changes in the syllabus or classroom materials can also occur as a result of students' performance on tests. In fact, especially in a large institution, formal tests serve a multitude of purposes.

2- Objective and Subjective Testing

Subjective and objective are terms used to refer to the scoring of tests an objective test has only one correct answer per item. A subjective test may result

in a range of possible answers, some of which are more acceptable than others. In every testing effort, we make decisions and choices under the influence of subjective judgments. Although great logical, psychometric, and empirical efforts have been made in the profession to avoid basing test construction on subjective views, it seems that subjectivity is an inevitable aspect of language test planning, construction, and validation.

In an oral test, for example, the testee must think of what to say and then express his/her ideas as well as possible; in a multiple-choice tests he has to weigh up carefully all the alternatives and select the best one. Furthermore, all tests are constructed subjectively by the tester: he/she decides which areas of language to test, how to test those particular areas, and what kind of items to use for his/her purpose. Thus, it is only the scoring of a test that can be described as objective. This means that a testee will score the same mark no matter which examiner marks the test. Since objective tests usually have only one correct answer, they can be scored mechanically. The fact that objective tests can be marked by computer is one important reason for their evident popularity among examining bodies responsible for testing large numbers of candidates.

Objective tests need not be confined to any one particular skill or element. Investigations, for example, have indicated that it is possible to measure writing ability by a series of objective tests. However, while it is misleading to associate examinations solely with writing and objective tests solely with grammar and vocabulary, certain areas of language may be tested more effectively at certain levels by one method rather than by another.

On the whole, objective tests require far more careful preparation than subjective examinations. In an examination, the examiner tends to spend a relatively short time on setting the questions but considerable time on marking.

In an objective test, the tester spends a great deal of time constructing each test item as carefully as he/she can, attempting to anticipate the various reactions of the testees at each stage. The effort is rewarded, however, in the ease of the marking.

The term examination and test are often used loosely. For our present purposes, however, the term *test* will generally be used to refer to a set of items that can be marked objectively, while examination will be used to refer to a set of longer subjective questions (for example. Oral interviews or compositions).

Objective tests are frequently criticised on the grounds that they are simpler to answer than subjective examinations. This is not always true as items, in an objective test, however, can be made just as easy or as difficult as the test constructor wishes. The fact that objective tests may generally look easier is no indication at all that they are easier. The constructor of a standardised achievement or proficiency test, not only selects and constructs the test items carefully, but analyses student performance on each item and rewrites where necessary so that the final version of the test discriminates widely. Setting the pass-mark, or the cutting-off point, may depend on the tester's subjective judgement or on a particular external situation. Objective tests (and, to a smaller degree, subjective tests) can be pre-tested before being administered on a wider basis; i.e. they are given to a small but truly representative sample of the test population, and each item is evaluated in the light of the testee's performance. This procedure enables the test constructor to calculate the approximate degree of difficulty of the test. Standards may then be compared not only between students from different areas or schools but also between students taking the test in different years. Another criticism is that objective tests of the multiple-choice type encourage guessing. However, four or five alternatives for each item are sufficient to reduce the possibility of guessing. Furthermore, experience shows

that candidates rarely make wild guesses: most base their guesses on partial knowledge.

A much wider sample of grammar, lexis and phonology can generally be included in an objective test than in a subjective test. In an essay, for example, the student can limit the range of the sentence structures and grammatical items he/she uses to those which he/she can produce correctly. However, the fact that a test is apparently completely objective gives no guarantee that it will be a good test. It will be a very poor test if:

- irrelevant areas are emphasised in the test simply because they are "testable";
- and the test items are poorly written.

It is true, that subjective judgments constitute a significant portion of any language testing endeavour. As Pilliner (1968) notes, there is no means to avoid subjective judgments even in tests which appear to be objectively designed, scored, and interpreted. Subjective judgments are so pervasive that they exist in all testing related areas (Alderson, 1993). Due to this ubiquitous impact, two equally knowledgeable testing authorities with different theoretical and/or practical backgrounds may come up with diametrically different decisions at each stage of testing. Moreover, each expert will probably be able to justify her/his decisions with self-corroborating "logical", "theoretical", or "empirical" evidence. Interestingly, each party can be bewildered if the other does not discern certain "straightforward facts".

If we come to accept the view that each theoretical position results from an amalgam of judgments and choices, we might begin to assume that the views held by each testing professional are relativistic and based on a torrent of previous judgments. Though this relativistic perspective has a long tradition, it might seem untenable from empirically-oriented scientific perspectives. Under

most circumstances, testing specialists in the same camp share many common beliefs as well as some differences of opinion. It is safe to say that, there are no two experts who think in identical ways about everything. This is partly due to the fact that there are no absolute criteria to base decisions on. As a result, after collecting bits of evidence, we still have to make decisions regarding which data to accept and which to reject. In sampling the kind of language, test content, subjects for the study, test tasks, test methods, and statistical procedures to determine reliability and validity and the like, we are bound to make subjective decisions for which, in the final analysis, there is no "definite evidence".

It is our belief that it should never be claimed that objective tests can do those tasks which they are not intended to do. They can never test ability to *communicate* in the target language nor can they evaluate actual performance. A good classroom test will probably contain both subjective and objective test items.

Conclusion

It is an accepted fact that teaching and testing are closely interrelated. Most of the reasons for testing take us back to teaching and the different people involved in an educational context, i.e. learners, teachers, parents or sponsors, course directors, the institution and society as a whole — in relation to formal test. This has to be taken into consideration in the elaboration of any test, be it part of summative or formative assessment, objective or subjective testing.

CHAPTER TWO

BASIC CONSIDERATIONS IN TEST DESIGN

Introduction	38
I-Developing a Test	38
1-Stages of Development of a Test	38
2-Test Items	40
3-Variables Affecting Test Performance	42
II- Criteria of Test Design	44
1 — Validity	45
1.1-Content Validity	46
1.2-Criterion Validity	51
1.3-cocnstruct Validity	56
1.4-Wash Back Validity	61
2 — Reliability	64
-Factors which Affect Reliability	65
3- Practicality	69
Conclusion	70

Introduction

The concepts of validity, reliability and practicality affect all aspects of test design, irrespective of the prevailing linguistic paradigm. In this chapter the nature of these key concepts is examined in detail. The status of the various types of validity, and how the concept of validity relates to those of practicality and reliability are highlighted.

I-Developing a Test

Although the actual designing of tests is in the hands of a small group of people working at headquarters, i.e., the group of course planners, classroom teachers have an active and essential participation in the process. This participation varies from analyzing first drafts of tests to completing feedback forms on the actual implementation and test results. Not to mention the process of standardization of marking

1-Stages of Development of a Test

When it comes to designing and implementing tests, fulfilling the requisites of reliability and validity plus satisfying the needs of all the participants involved in the educational process requires a very sophisticated operation. The complexity of the whole operation can be explained by the level of detail and care with which each stage has to be conducted. The testing operation comprises these four stages; with each one feeding into the next one:

- test design ;
- test implementation ;
- test marking ;

- and test validation or analysis: evaluation stage.

The evaluation stage comprises statistical item analysis as well as evaluation of each test as a whole, on the basis of feedback produced by teachers.

The stage of test design, for instance, has to involve the scripting and recording of listening texts, the production of suitable illustrations and a very special care in relation to rubrics and lay-out. Not to mention, the concerns with construct and content validity.

One of the aims of the evaluation stage in the process of test design is the creation of a bank of validated items. However, several factors make it difficult for a large enough bank of test items to be stored and re-used, so that the test design phase can be made easier. In the first place, there is the problem of considering test item in isolation, i.e. separated from the context of the whole test where they were inserted, thus spoiling a certain internal coherence and balance of a test.

Secondly, the more a test is related to syllabus or set of materials, the more difficult it is to use it in different contexts. Different syllabuses, thus, would require new tests.

Thirdly, if tests include "authentic" or "simulated authentic" materials, these will often be related to current affairs which get outdated fairly easily.

Fourthly, the process of validation is by nature a dynamic one, with some hypothesis being disconfirmed and new hypotheses being created all the time. Hence the need for new testing instruments.

Given the multitude of aspects and interests involved in it, such a testing operation is not only complex but extremely time consuming. Not enough time is left, then, for other areas such as monitoring of continuous assessment, or even for piloting tests before their actual implementation.

2- Test Items

The actual question of what is to be included in a test is often difficult because a mastery of skills is being assessed rather than areas of knowledge. Although the construction of a language test at the end of the first or second year of learning English is relatively easy if we are familiar with the syllabus covered, the construction of a test at a fairly advanced level where the syllabus is not clearly defined is much more difficult.

The longer the test, the more reliable a measuring instrument will be (although length, itself, is no guarantee of a good test). Few students would want to spend several hours being tested — and this would be undesirable both for the tester and the testee. But the construction of short tests which function efficiently is often a difficult matter. Sampling is of paramount importance. The test must cover an adequate and representative section of those areas and skills it is desired to test.

If all the students who take the test have followed the same programme, we can simply choose areas from this programme, seeking to maintain a careful balance between tense forms, prepositions, articles, lexical items, etc. Above all, the kind of language to be tested would be the language used in the classroom and in the student's immediate surroundings or the language required for the school or the work for which the student is being assessed.

The problem arises when some testees may have attended different schools or followed different courses. In such a case, the testees will all experience problems of a similar nature as a result of the interference of their first- language habits. These problem areas are to be identified and included in a cross-section in the test, particularly in those sections of the test concerned with the language elements.

Before starting to write any test items, the test constructor should draw up a detailed table of specifications showing aspects of the skills being tested and giving a comprehensive coverage of the specific language elements to be included. A classroom test should be closely related to the ground covered in the class teaching, an attempt being made to relate the different areas covered in the test to the length of time spent on teaching those areas in class. There is a constant risk of concentrating too much on testing those areas and skills which most easily lend themselves to being tested. It may be helpful for the teacher to draw up a rough inventory of items according to importance. For example, a teacher wishing to construct a test of oral expression might start by examining the relative weighting to be given to the various areas in the light of the teaching that has just taken place: say, pronunciation and fluency (40%), stress and intonation (15%), grammar (25%), and meaning (20%). It must be emphasised that this list is merely one example of the kind of inventory which can be drawn up beforehand, and it is not intended to represent a particular set of priorities. on the contrary, the scale is left to the appreciation of the teacher who will act according to the objectives of the test.)

It is our strong belief that a good test should never be constructed in such a way as to trap or mislead the testee into giving an incorrect answer. When techniques of error analysis are used, the setting of deliberate traps or pitfalls for

the unwary student should be avoided. Many testers, themselves, are caught out by constructing test items which succeed only in trapping the more able student. In the following example of a multiple choice question, we found out that the testee has to select the correct answer © but the whole item is constructed so as to trap the testee into making choice B or D. When this item actually appeared in a test, it was found that the more proficient students, in fact, chose B and D, as they had developed the correct habit of associating the tense forms *have see* and *have been seeing* with *since and for* chose the "correct answer".

To summarise, we can say that all tests should be constructed primarily with the intention of finding out what students know — not trapping them. By attempting to construct effective language tests, the teacher can gain a deeper insight into the language being tested and the language learning processes involved.

3-Variables Affecting Test Performance.

The investigation of the degree of accuracy with which a testing instrument measures language abilities makes one look at trait and method factors, and the way in which they "contaminate" one another.

Trait-or ability- factors represent the knowledge that a test claims to measure, for example, reading, listening to train announcements, producing accurate sounds. Method factors, on the other hand, have to do with the way in which the display of knowledge is prompted and assessed. In this sense, one could say that the less the interference of method factors, the better a test will be as a measure of language abilities. This, obviously, is more easily said than done but, it should nevertheless be a true concern for test designers.

Skehan (1988) presents three categories of trait/method influences on test scores: language based problems, learner based problems and method factors.

- **Language- based problems** relate to variabilities of speech which speakers find themselves in. They can be a result of:
 - e- linguistic context, i.e., influences of the immediate linguistic environment on the production of syntactic forms, as well as the discourse environment which may affect style and accuracy of production;
 - r social situation, in which different degrees of formality or demands on a particular image to be projected by the speaker may cause variations from very careful to casual speech;
 - + - interlocutor relationship: this is related to the social situation outlined above where perception concerning difference from planned to "spontaneous" discourse;
 - r amount of context, in which the degree of context-embedded ness and familiarity with the topic may affect communication.
- **Learner-based problems:** these include background variables such as sex, age mother-tongue, cognitive style, the role of intelligence and affective reactions to testing.
- **Method factors** have to do with test formats and item types, and the extent to which they might enhance or restrict language production.

Some of the factors outlined above are particularly important when one is dealing with young beginners. Young learners' production may be particularly affected by lack of familiarity with both topic and question format, a high level of abstraction or of decontextualized language manipulation; emotional reactions to testing situations if they have not yet been used to being formally assessed or if the regular school system places too much emphasis on testing, thus leading to a high level of anxiety.

A programme of test design and implementation then should be as broadly and carefully planned as possible to account for all these different variables. From this perspective, test research and statistical measures of reliability are important for the sake of accuracy of score interpretation and mainly as a response to an educational commitment.

II-Criteria of Test Design

As pointed out by Milanovic (1988), however, the average classroom teacher still knows very little about testing. Most teacher training courses either do not include the subject altogether or treat it in a very brief, superficial fashion. We believe that this may be a result of the fact that tests are usually regarded as something alien to the daily classroom environment, something which is imposed on teachers and learners by the institution.

Testing systems usually receive relatively little attention from people working in the field of education. This situation is also true for English as a Foreign Language (EFL) teaching, although the last decade has witnessed a great deal of research and debate on language tests and the assessment of language proficiency in general.

1-Validity

Validity means the truth of the test in relation to what it is supposed to evaluate. It concerns the relevance and usefulness of what we are measuring. The difficulty in setting such tests lies in the problem of how sure we can be about what *is* actually being measured. Whether it is consistent with the worthwhile quality we think we are measuring. The concept of validity (does the test measure what it is intended to measure) refers to the degree to which a test measures what it is intended to measure and whether appropriate inferences can be drawn from test results. It is judged from many types of evidence including the consequences of translating test-based inferences into decisions or policies that can affect individuals or institutions. Tests must demonstrate an acceptable level of validity before they are used to make decisions.

When discussing test validity, as pointed out by Bachman (1990) it is useful to clarify some key points. Firstly, validity is not an intrinsic quality that some tests possess and others do not. The process of validation has to do with the interpretation and use of the information gathered through a testing instrument. In other words, the validity of a test has to be measured against a set of criteria in relation to the abilities being tested and to what is to be made of test scores. Secondly, as mentioned earlier in this chapter, reliability is a basic requirement for validity, and that writers should try to reconcile the two notions.

Thirdly, the process of validation comprises complementary types of evidence, which can be grouped into three general types: content relevance (or content validity); criterion relatedness (or criterion validity); and meaningfulness of construct (or construct validity) (Messick 1980; 1988). However, more recently, validity has been viewed as a sub piece of information that includes aspects of the three types of validity. In addition to these types of validity,

several authors (Porter 1983, Morrow 1986, Wall 1993) referred to another type of validity- washback validity.

1.1- Content Validity

Perhaps the first aspect which learners try to identify in a test is the extent to which the test adequately covers the syllabus they have been through, thus giving them a chance to show how much they have learnt. The aspect of content validity has to do with the representative ness of the samples of a syllabus which can be included in a test. In other words, given that no test can possibly cover all aspects of a syllabus —not to mention what has been learnt beyond the syllabus- it is important that a test contains a representative sample of points perceived to be essential within a given syllabus. The aspect of content validity will obviously be a reflection of the degree of clarity and comprehensiveness of course objectives, syllabus specification of question- types.

In relation to progress and achievement tests, especially when young learners are involved, the process of content validation will also take into account a balance which has to be achieved between familiarity with topics, subject matter and question types and ability to abstract and apply a given knowledge to situations which are new, to a certain extent. In other words, if it is true that the test should sample a given syllabus, at the same time one should make sure those learners are asked to apply their knowledge of the language and not simply memorize exercises done in class.

Content Validity is then evaluated by showing how well the content of the test samples the subject matter about which conclusions are to be drawn. It is established through a careful examination of test items as they relate to a specified content area often specified by instructional objectives. The greater the

match between instructional objectives and the test items, the greater the content validity. This is especially important in evaluating achievement tests.

Anastasi (1982:131) defined content validity as

... essentially the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured.

She provided a set of useful guidelines for establishing content validity:

the behavior domain to be tested must be systematically analyzed to make certain that all major aspects are covered by the test items, and in the correct proportions;

.....
the domain under consideration should be fully described in advance, rather than being defined after the test been prepared;

.....
content validity depends on the relevance of the individual 's test responses to the behavior area under consideration, rather than on the apparent relevance of item content. (Anastasi ,op-cit p: 132)

It is said that the more a test simulates the dimensions of observable performance and accords with what is known about that performance, the more likely it is to have content and (construct) validity. We can often only talk about the communicative construct in descriptive terms and, as a result, we become involved in questions of content relevance. Thus, for Kelly (1978:8) content

validity seemed "an almost completely overlapping concept" with construct validity, and for Moller (1982:68). . . the distinction between construct and content validity in language testing is not always very marked, particularly for tests of general language proficiency.

Given the restrictions on the time and resources available to those involved in test construction, especially for use in the classroom, it is often only feasible to focus on the a priori validation of test tasks. In these cases, particular attention must be paid to content validity in an attempt to ensure that the sample of activities to be included in a test is as representative of the target domain as is possible.

A primary purpose of many communicative tests is to provide a profile of the student's proficiency, indicating in broad terms the particular modes where deficiencies lie. Content validity is concerned with the extent to which the selection of test tasks is representative of the larger universe of tasks of which the test is assumed to be a sample (Bachman et al.1981).The directness of fit and adequacy of the test sample is thus dependent on the quality of the description of the target language behavior being tested. Establishing content validity is problematic, given the difficulty in characterizing language proficiency with sufficient precision to ensure the representativeness of the sample of tasks included in a test. Additional threats to validity may arise out of attempts to operationalise real-life behavior in tests, especially where some sort of quantification is necessary, either in the task or the method of assessment. These difficulties do not, however, absolve the test constructor from attempting to make tests as relevant in terms of content as possible.

Further, there is a need to look closely at test specifications to make sure that they describe adequately what ought to be tested. A close scrutiny of the

specification for a proficiency test by experts in the field (or colleagues in the case of classroom achievement tests) and the relating of the specification to the test as it appears in its final form is essential (Weir 1983). This would provide useful information as to what the test designer intends to test and how successful the item writers had been in implementing the specification in the test realization. Mere inspection of the modules in the test, even by language and subject experts, does not necessarily guarantee the identification of the processes actually used by candidates in taking them. In addition, it would be valuable to employ ethnographic procedures to establish the validity of items in a test. A useful procedure is to have a small sample of the test population introspect on the internal processes that are taking place in their completion of the test items. This would provide a valuable check on experts' surface-level judgments on what was being tested and would contribute to the establishment of guidelines for the conduct of this type of methodological procedure in future investigations of test validity.

It is crucial for a test supposedly based on specified enabling skills to establish that it conforms to the specification, especially if claims are made for these being representative of the domain in question. To the extent that the content is made explicit, the concern also becomes one of face validity which Porter (1983) describes as perhaps the most contentious validity that might be invoked.

Lado (1961), Davies (1965), Ingram (1977), Pamler (1981) and Bachman et Al. (1981) have all discounted the value of face validity. Bachman et Al (1981: 55) argues as follows:

Since there is no generally accepted procedure for determining whether or not a test demonstrates this

characteristic, and since 'it is not an acceptable basis for interpretative inferences from test scores', we feel it has no place in the discussion of test validity.

Anastasi (1982:136) pointed out that face validity

...is not validity in the technical sense; it refers, not to what the test actually measures, but to what it appears superficially to measure. Face validity pertains to whether the test 'looks valid' to the examinees who take it, the administrative personnel who decide on its use, and other technically untrained observers.

If a test does not have face validity though, it may not be acceptable to the students taking it, or the teachers and receiving institutions who may make use of it. If the students do not accept it as valid, their adverse reaction to it may mean that they do not perform in a way which truly reflects their ability. Again Anastasi (1982: 136) says:

Certainly if the test content appears irrelevant, inappropriate, silly or childish, the result will be poor cooperation, regardless of the actual validity of the test. Especially in adult testing, it is not sufficient for a test to be objectively valid. It also needs face validity to function effectively in practical situations.

1.2-Criterion Validity

According to Bachman (1990:288) "Criterion validity is an indication of the relationship between test scores and some criterion which we believe is also an indicator of the ability tested". This is concurrent validity. The other aspect of criterion validity relates to the extent to which a test can adequately predict some future behaviour. This is called predictive validity and requires an examination of test scores and actual performance on a job or course.

Attempts at establishing concurrent validity usually take two forms. One of them is an investigation of the differences in performance on the same test by different groups of test-takers perceived as having different levels of ability. Validation then is related to the degree of discrimination which the test allows. The other more typical form of concurrent validity is an indication of the degree of correlation between two tests. A word of caution can be said in this respect: that concurrent validation requires scores on two different measures of the same ability. The problem then lies in determining whether the two different tests do really measure the same aspects. One way in which test designers try to cope with the problem is by correlating "internal" exams with some «external" measure, such as public exams, for example the TOEFL or IELTS .However, in the case of progress (or even achievement) tests for young beginners, it is very difficult to find equivalent tests which have been trialed on a large scale. This sort of correlation can be done, in this case, by comparing test scores to teachers' continuous assessment of learners' performance. Some studies have indicated indeed that teachers' ratings are very good ways of establishing concurrent validity (Ingram, 1974; and North, 1990).

As for predictive validity, in the context of a school system with different levels and promotion based on test scores, a study of the predictive validity of a test would involve the correlation of scores on progress tests and performance on the subsequent level and corresponding progress test. This is something

which most teachers do on an informal basis, usually to try and spot "weak" and "good" students at the beginning of each term and observe how they develop. The problem with predictive validity, in this case, is the individuals' change in the way they approach a course of instruction from one term to the other. More or less effort or opportunities to extend and consolidate the work done in the classroom may then lead to differences in performance, irrespective of scores on previous tests.

Criterion related validity is evaluated by comparing the test scores with one or more external variables (called criteria) considered to provide a direct measure of the characteristic or behaviour in question. There are two subtypes: predictive and concurrent validity. Predictive validity indicates the extent to which an individual's future level on the criterion is predicted from the test performance. For example, a student's score on a third grade speaking test might be used to predict the final grade a student would receive in fourth grade speaking (the final speaking grade is the criterion). Concurrent validity indicates the extent to which the test score estimates an individual's present standing on the criterion. For example, a student's score on a speaking test might be correlated with the speaking grade received during that particular grading period. In all cases, criterion-related validity is represented by the correlation computed between the test score and the criterion measure. This is a predominantly a posteriori concept, concerned with the extent to which test scores correlate with a suitable external criterion of performance: what Ingram, 1974:18) termed 'pragmatic validity'. Criterion-related validity divides into two types (Davies 1977), concurrent validity, where the test scores are correlated with another measure of performance, usually an older established test, taken at the same time (Kelly 1978 ; Davies 1983) and predictive validity, where test scores are correlated with some future criterion of performance (Bachman et al. 1981)

For many authorities, external validation based on data is always superior to the 'armchair speculation of content validity'. Davies (1983:1) has argued forcefully that external validation based on data is always to be preferred:

The external criterion, however hard to find and however difficult to operationalise and quantify remains the best evidence of a test's validity. All other evidence, including reliability and the internal validities is essentially circular.

He quotes Anastasi op-cit: 146 on the need for independently gathered external data: Internal analysis of the test, through item-test correlations, factorial analysis of test items, etc. Is never an adequate substitute for external validation.

Though this concept of criterion-related validity is more in keeping with the demands of an empiricist-operationalist approach, the problem remains that a test can be valid in this way without our necessarily knowing what the test is measuring, i.e., it tells us nothing directly about its construct validity. Morrow (1979:147) drew attention to the essential circularity of employing these types of validity in support of a test:

Starting from a certain set of assumptions about the nature of language and language learning will lead to language tests which are perfectly valid in terms of these assumptions whose value must inevitably be called into question if the basic assumptions themselves are challenged.

For Jakobovits (1970:75), the very possibility of being able to construct even one communicative test appeared problematic:

...the question of what it is to know a language is not well understood and, consequently, the language proficiency tests now available and universally used are inadequate because the attempt to measure something that has not been well defined.

Even if it were possible to construct a valid communicative test, there would still be problem in establishing sufficiently valid criterion measures against which to correlate it. Hawkey (1982:153) felt this to be particularly problematic for tests conceived within a communicative paradigm:

At this developmental stage in communicative testing, other tests available as criteria for concurrent validation are likely to be less integrative/communicative in construct and format and thus not valid as references for direct comparison.

There is a distinct danger that one might be forced to place one's faith in a criterion measure which may in itself not be a valid measure of the construct in question. One cannot claim that a test has criterion-related validity because it correlates highly with another test, if the other test itself does not measure the criterion in question. In this sense, it seems pointless to validate a test conceived within the communicative paradigm against tests resulting from earlier paradigms if they are premised on such different constructs. Similarly if equivalent but less efficient tests are not available against which to correlate,

other external non-test criteria might need to be established. Establishing these non-test criteria for validating communicative tests could well be problematic. Even if one had faith in the external criterion selected, for example, a sample of real life behaviour, the quantification process which it might be necessary to subject this behaviour to in order for it to become operational might negate its earlier direct validity.

Though caution is advocated in the interpretation of these criterion-related validity measures, they are still considered to be potentially useful concepts. For example, one might be very wary of tests that produced results seriously at variance with those of other tests measuring the same trait, especially if the latter had been found to have construct validity. It is particularly important not to try to establish criterion-related validity for a test through empirical monitoring whenever the candidates' futures may be affected by its results. For example, given the variety of language qualifications currently acceptable as evidence of language proficiency for entry into tertiary-level study in Britain (Weir 1983), there is some cause for concern about the equivalence of such a broad spectrum of tests. Where is the empirical evidence for the equivalence of one entry qualification with another? In the case of predictive validity, it may be that in certain circumstances the predictive power of the test is all that is of interest. If all one wants is to make certain predictions about future performance on the basis of the test results, this might entail a radically different test from that where the interest is in providing information to allow effective remedial action to be taken. If predictions made on the basis of the test are reasonably accurate then the nature of the test items and their content might not be more important.

1.3— Construct Validity

Construct validity has to do with the psychological properties (Palmer and Groot, 1981) or abilities being tested — in other words trait factors. It relates then to the process of validating a test, not against another test, but against a framework or theory of "knowledge".

With an English Foreign Language (EFL) test, when attempting to establish construct validation, one would be looking for ways of defining, for example, "knowledge of grammar", "ability to converse", «ability to infer meaning of unknown words",

Palmer and Groot (op. Cit: 4-5) identify five steps of a general procedure for investigating construct validity:

- defining the traits one is trying to measure ;
- selecting existing tests or writing new tests to operationalized defined traits ;
- stating hypotheses about relationships between scores on various test ;
- administering and scoring tests ;
- and comparing results with formulate hypotheses.

Talking about constructs (or abilities) being tested, one cannot **help** but consider the way in which these constructs may be operationalized, In other words, the relationship between trait and method, or between the abilities being measured and the question-types being used to prompt students to demonstrate such abilities. An example could be a test which aims at measuring learners' abilities to communicate orally. A question which requires learners to share information with friends, telling each other about an event which happened over the previous weekend would possibly be a good way of testing the ability to converse.

Construct Validity is evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain explanatory concepts or constructs account for performance on the test. Common examples of constructs include intelligence, language proficiency, aptitude and self-concept. Construct validity is established both logically and empirically. For example, a test which intends to measure English language proficiency would be examined to be sure that it included items that measured oral language skills (listening and speaking) as well as literacy skills (reading and writing). In addition, studies would be conducted to determine if the tests were sensitive to changes occurring as the result of instruction designed to improve English proficiency. Studies also would be conducted to determine if the test correlated with other measures of English proficiency.

The concept of construct validity can be interpreted in a number of ways in the literature. For Cronbach (1971:443), "Construction of a test itself starts from a theory about behavior mental organization derived from prior research that suggests the ground plan for the test."

Kelly (1978:8) supported this view, commenting that:

... the systematic development of tests requires some theory, even an informal, inexplicit one, to guide the initial selection of item content and the division of the domain of interest into appropriate Sub-areas.

For Anastasi (1982: 144) defined it as

. . . the extent to which the test may be said to measure a theoretic construct or trait...Each construct is developed to explain and organize observed response consistencies. It derives from established inter-relationships among behavioral measures... Focusing on a broader, more enduring and more abstract kind of behavioral description... construct validation requires the gradual accumulation of information from a variety of sources. Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill.

She also specified that:

...content, criterion-related and construct validation do not correspond to distinct or logically co-ordinate categories. On the contrary, construct validity is a comprehensive concept which includes the other types. (Anastasi ,op-cit1982: 153)

She argued that the theoretical construct, trait or behaviour domain measured by any test can be defined in terms of the operations performed in establishing the validity of the test.

The view expressed below differs only insofar as external empirical data are seen as a necessary but not a sufficient condition for establishing the adequacy of a test for the purpose for which it was intended. Though there is a lack of an adequate theoretical framework for the construction of communicative tests, this does not absolve test constructors from trying to establish a priori construct validity for a test conceived within the communicative paradigm. A test should always be designed on a principled basis, however limited the underlying theory, and, wherever possible after its administration, statistical validation procedures should be applied to the results to determine how successful the test has been in measuring what it intended to measure.

In the past, little attention was accorded to the non-statistical aspects of construct validity. In the earlier psychometric-structuralist approach to language testing the prevailing theoretical paradigm lent itself easily to testing discrete elements of the target language, and little need was seen for much a priori deliberation on the match between theory and test. Additionally, the empiricism and operationalism of those working in educational measurement made the idea of working with non-objective criteria unattractive. The notions of concurrent and predictive validity, more consistent with the principles of operationalism and the desire for an objective external criterion, took precedence. Now, construct validity is viewed from a purely statistical perspective in much of the recent American literature (Palmer et al. 1981; Bachman et al.1981). It is seen principally as a matter of the a posteriori statistical validation of whether a test has measured a construct which has a reality independent of other constructs. The concern is much more with the a posteriori relationship between a test and the psychological abilities, traits and constructs it has measured than with what should have been elicited in the first place.

To establish the construct validity of a test statistically, it is necessary to show that it correlates highly with indices of behavior that one might theoretically expect it to correlate with, and that it does not correlate significantly with variables that one would not expect it to correlate with. An interesting procedure for investigating this is the convergent and discriminant validation process first outlined by Campbell and Fiske (1959) and later used by Bachman and Palmer (1981). The latter argue that the strong effect of test method that they discovered, points to the necessity of employing a multi-trait multi-method matrix as research paradigm in construct validation studies about this aspect. The experimental design of the multi-trait/multi-method matrix has been criticized especially in relation to more direct tests of language proficiency, but nevertheless it deserves further empirical investigation, as very few studies have been reported.

In contrast to this emphasis on a posteriori statistical validation, there is a body of opinion which holds that there is an equally important need for construct validation at the a priori stage of test design and implementation. It would seem self-evident that the more fully we are able to describe the theoretical construct we are attempting to measure, at the a priori stage, the more meaningful might be the statistical procedures contributing to construct validation that can subsequently be applied to the results of the test. Statistical data do not in themselves generate conceptual labels. We can never escape from the need to provide clear statements concerning what is being measured, just as we are obliged to investigate how adequate a test is in operation, through available statistical procedures.

Construct validation seems to involve at least two problems. The first one is the assumption that abilities or constructs can be precisely and finely defined and divisible into components which can be measured separately. The second

problem is that the whole process tends to be cyclical since theories about language lead to the design of instruments which are observed through the point of view of the same theory. If hypotheses are not confirmed, new theories will be formulated and the cycle will continue. This does not mean, however, that the process is useless. Indeed that is the way in which science seems to develop. The process can in fact generate valuable insights, apart from being a way of ensuring clarity of objectives.

Construct validity is in fact the most important aspect of the process of validation, since it must be theory driven and conducive with a framework of language ability. It could be said that content validity will be concerned with the extent to which certain constructs are operationalized through the use of grammar, lexis, topics or tasks and concurrent validity, should be subordinated to construct validity; if that is not the case, the mere correlation between two tests tends to be pointless

1.4- Wash back Validity

Shohamy, Donitsa-Schmidt and Ferman (1996: 298) defined washback as "the connections between testing and learning". Gates (1995: 101) defined washback simply as "the influence of testing on teaching and learning". Messick (1996: 241) referred to washback as ". . . the extent to which the introduction and use of a test influences language teachers and learners to do things they would not otherwise do that promote or inhibit language learning". The washback effect clearly has to do with the effect of testing on the teaching and learning processes.

In other words, the washback effect can be defined as the direct or indirect effect of examinations on teaching methods. According to the effect of

examinations on what we do in the classroom, we may refer to 'positive' and 'negative' backwash (Heaton 1990: 170. Hughes 1989: 1). Although washback is an important factor in classrooms wherever examinations play a dominant role in the educational process, it has not been fully explored. It is not mentioned in the index to such Standard English Language Teaching (ELT) handbooks as Stern (1983), Howatt (1984), or Harmer (1991), and reference books such as Richards *et al.* (1985) and Seaton (1982) do not consider it worthy of an entry. Heaton (1990) and Hughes (1989) discuss it, rather sketchily; referring to it as 'overt' backwash. Alderson and Wall (1993), suggest that washback is more complex than has hitherto been assumed. They make the point that there is no one-to-one relationship between tests, good or bad, and their effect on the classroom. In their view, before a test has any impact on classroom practice, it is mediated by factors such as the place of examinations in particular, societies, the teacher's competence, and the resources available within the school system.

Whether the washback effect is positive or negative, how it operates in particular contexts —indeed, whether it exists at all—must be explored empirically. Many of the assumptions about backwash are untested and simplistic. Alderson and Wall (1993) point out that very little observation of the effect has been carried out, and that what evidence there are points to the highly complex nature of the process. Professional neglect of the washback effect (what it is, how it operates, and its consequences) is one of the main reasons why new methods often fail to take root in language classes. Many teachers, trapped in an examination preparation cycle, feel that communicative and humanistic methodologies are luxuries they cannot afford. When the market calls on teachers and institutions to produce quantifiable results, it usually means good *examination* results. Sound teaching practices are often sacrificed in an anxious attempt to 'cover' the examination syllabus, and to keep ahead of the

competition. In summary, 'negative washback', as experienced by the learner, means language learning in a stressful, textbook-bound environment.

The difficulties of precisely determining what it is that needs to be measured perhaps argues for a greater concern with what has recently been termed 'washback validity' (Morrow 1986) or more commonly the washback of the test on the teaching and learning that precedes it. (Porter 1983 and Weir 1983) .Given that language teachers operating in a communicative framework normally attempt to equip students with skills that are judged to be relevant to present for future needs and to the extent that tests are designed to reflect these, the closer the relationship between the test and the teaching that precedes it, the more the test is likely to have construct validity. We can say that a communicative approach to language teaching is more likely to be adopted when the test at the end of a course of instruction is itself communicative. A test can be a very powerful instrument for effecting change in the language curriculum as shown in recent developments in the field of language testing. A suitable criterion for judging communicative tests might well be the degree to which they satisfy students, teachers and future users of test results, as judged by some systematic attempt to gather quantifiable data on the perceived validity of the test.

If the test passes the first a priori validity handle, it is then worthwhile establishing its validity against external criteria, through confirmatory a posteriori statistical analysis. If the first stage, with its emphasis on construct, content, face and washback validities, is bypassed, then we should not be too surprised if the type of test available for external validation procedures does not suit the purpose for which it was intended.

In searching the literature on washback we found that Hughes (1989), Heyneman and Ransom (1990), Shohamy (1992). Kellaghan and Greaney

The degree of inter-marker reliability is established by correlating the scores obtained by candidates from marker A with those from marker B. The consistency of each individual marker (inter-marker reliability) is established by getting them to remark a selection of scripts at a later date and correlating the marks given on the two occasions. (Anastasia 1982 for a clear and accessible introduction to the necessary statistics; also of use are Crocker, (1981) and more recently Woods et al. (1986). The concern of the tester is how to enhance the agreement between markers by establishing, and maintaining adherence to explicit guidelines for the conduct of this marking. The criteria of assessment need to be established and agreed upon, and markers need to be trained in the application of these criteria through rigorous standardization procedures (Murphy 1979). During the marking of scripts, there needs to be a degree of cross-checking to ensure that agreed standards are being maintained.

Other aspects which concern test administration include the adequacy of premises and equipment used; the time allowed for students to answer the different papers; the amount of information to be provided —or not- by test administrators; aspects of security; amount of preparation or training in answering different parts of a test; clarity of instructions or rubrics and of examples provided.

Individual factors include aspects which relate to test takers' cultural/educational background or to psychological aspects such as: anxiety, fatigue, degree of familiarity with test-type, cognitive style, previous experiences in test-taking or empathy with test administrators or examiners. Probably the most difficult task for test planners in relation to test reliability is to anticipate variability in terms of test takers. Although these aspects are taken into consideration at the stage of test design, i.e., before the test is actually

implemented, it is important that at a post-test stage ,the influence of individual factors on test scores be investigated.

Reliability refers to:

If tests are scored by human judges, to what extent do different judges agree in their estimation of student achievement?

How consistently a test measures whatever it is measuring and how generalizable the results are?

Will a student's score today be close (if not identical) to his/her score tomorrow?

Reliability can be grouped into three general types: test-retest, where the same test is readministered to the same sample population after a short intervening period of time, internal consistency obtained through dividing a test into two halves (for example, one-half composed of the even numbered items and one-half of the odd numbered items) and then correlating the scores on each half of the test. This is called split-half reliability, and equivalent forms.

The third aspect of reliability that of parallel- forms reliability, is often very difficult to achieve for both theoretical and practical reasons. To achieve it, two alternative versions of a test need to be produced which are in effect clones of each other. The reliability of the versions is directly proportional to the similarity of the versions obtained when administered to the same test population.

Equivalent forms reliability is determined by administering two different forms of a test to the same group of people and then correlating the two scores to determine the reliability index. As little time as possible (for example,

one day) should elapse between the two administrations. This is a measure of the equivalence of the test scores. If more time is allowed to elapse (for example, longer than one or two days), the reliability index becomes a measure of both stability and equivalence, or a combination of test-retest and equivalent forms reliability.

3-Practicality

The practicality of a test refers to the degree to which it is cost effective and easy to administer. The number of testees, the time constraints for testing and marking, and the available human and physical resources need to be considered carefully before an assessment scheme is chosen. This is not only an issue of money, but also of the perceptions of those who will be taking and using the test. If a test can be administered efficiently by assessors and markers, this increases the validity and reliability of the results as a whole. On the other hand, a valid and reliable test is of little use if it does not prove to be a practical one. This involves economy, ease of administration, scoring, and interpretation of results. The longer it takes to construct, administer and score, and the more skilled the personnel and equipment that are involved in the test, the higher the costs are likely to be.

The duration of the test may affect its successful operation in other ways. For example, a fatigue effect on the candidates, administrative factors such as staff to invigilate and the availability of rooms in which to sit the examination; all have to be taken into consideration. It is thus highly recommended to make the test as short as possible, consistent with the need to meet the validity and reliability and practicality criteria referred to above.

Practicality, in the sense of financial viability, may indeed prove to be the real obstacle in the way of the development of communicative tests. Tests of this type are difficult and time-consuming to construct, require more resources to administer, demand careful training and standardization of examiners and are more complex and costly to mark and report results on. The increased per-capita cost of using communicative tests in large-scale testing operations may severely restrict their use.

Conclusion

However problematic, we believe that there is clearly an imperative need to try and develop test formats and evaluation criteria that provide the best overall balance among reliability, validity and practicality in the assessment of communicative skills.

To provide profiles rather than standard scores, each part of the profile will need to reach an acceptable degree of reliability. To achieve satisfactory reliability, communicative tests may have to be longer and have multiple scoring. To enhance validity by catering for specific needs and profiling, several tests will be needed, thus further raising the costs as compared to those of single general tests available for large populations.

All the considerations to be made in order to ensure the good quality of tests as instruments of evaluation, call for a complex process. It must be conducted very carefully and will probably require a good deal of investment in terms of time and personnel, especially in a large institution, which is the case for the Department of English under study.

CHAPTER THREE

APPROACHES TO LANGUAGE TESTING

Introduction	72
I- The Structuralist Approach	73
II- The Psycho- Sociolinguistic Approach	77
III- The Communicative Approach	82
1-Communicative Language Ability	82
2-Characteristics of a Language Test	87
Conclusion	97

Introduction

In this chapter, we will discuss the different approaches to language testing and their limitations in terms of the criteria of validity, reliability and practicality. These approaches to testing can be usefully described in terms of the particular focus they represent. Although these approaches are listed here in chronological order, they should not be considered as being strictly confined to certain periods in the development of language testing. Nor are the four approaches always mutually exclusive. It is our belief that a useful test will generally incorporate features of several of these approaches. A test may have certain inherent weaknesses simply because it is limited to one approach, however attractive that approach may appear.

Davies (1978:149) argued that by the mid '70's, approaches to testing seemed to fall along a continuum which stretched from 'discrete' item tests at one end to integrative tests such as cloze at the other. He took the view that in testing, as in teaching, there was a tension between the analytical approach on the one hand and the integrative approach on the other, and considered that "... the most satisfactory view of language testing and the most useful kinds of language tests, are a combination of these two views, the analytical and the integrative." He explains that it was probable, in any case, that no test could be wholly analytical or integrative and that:

The two poles of analysis and integration are similar to (and may be closely related to) the concepts of reliability and validity. Test reliability is increased by adding to the stock of discrete items in a test; the smaller the bits and the more of these there are, the higher the potential reliability. Validity however is

increased by making the test truer to life, in this case more like language in use. Davies, op-cit: 149

Oiler (1979), on the other hand, believes that testing should focus on the integrative end of the continuum away from the 'discrete point' approach to testing, to what he termed "the psycholinguistic-sociolinguistic era": the age of the integrative test.

Madsen (1983) describes that language testing as having evolved through three major stages, which reflect people's attitudes towards the goals of language teaching and language learning:

- the Intuitive Stage which focuses on subjective testing and is dependent on personal impressions of the teachers,
- the Scientific Stage which stresses objective evaluation focusing on language usage.
- the Communicative Stage which emphasizes evaluation of language use rather than usage

I- The Structuralist Approach

This approach is characterised by the view that language learning is chiefly concerned with the systematic acquisition of a set of habits. It draws on the work of structural linguistics, in particular the importance of contrastive analysis and the need to identify and measure the learner's mastery of the separate elements of the target language: phonology, vocabulary and grammar. Such mastery is tested using words and sentences completely divorced from any context, on the grounds that a larger sample of language forms can be covered in the test in a comparatively short time. The skills of listening, speaking, reading and writing are also separated from one another as much as possible because it is considered

essential to test one aspect at a time. Such features of the structuralist approach are valid for certain types of test and for certain purposes; for example, the desire to concentrate on the testees' ability to write by attempting to separate a composition test from reading (i.e. by making it wholly independent of the ability to read long and complicated instructions or verbal stimuli) is recommendable in certain respects.

This approach to measurement emphasises on reliability and objectivity. Structuralists have been able to show clearly that such traditional examinations as essay writing are highly subjective and unreliable, as a result the need for statistical measures and validity is of the utmost importance in testing. The clear advantages of testing 'discrete' linguistic points are that they yield data which are easily quantifiable, as well as allowing a wide coverage of items. Tests which focus on 'discrete' linguistic items are efficient and have the usual reliability of marking associated with objectively scored tests,

The limitation of the 'discrete-point' approach and the various formats employed in it is that they suffer from the defects of the construct they seek to measure. According to Oller (op-cit: 212)

Discrete point analysis necessarily breaks the elements of language apart and tries to teach them (or test them) separately with little or no attention to the way those elements interact in a larger context of communication. What makes it ineffective as a basis for teaching or testing languages is that crucial properties of language are lost when its elements are separated. The fact is that in any system where the parts interact to produce properties and qualities that do not exist in the part separately, the whole is greater than the sum of its parts

... organizational constraints themselves become crucial properties of the system which simply cannot be found in the parts separately.

From Oiler's point of view, we can say that testing a candidate's linguistic competence is a necessary, but not sufficient component of a test battery. In real life, for example, people taking a driving test are required to demonstrate that they perform the task; the licensing authority does not depend solely on a pencil and paper test to inform about the extent of testees' knowledge concerning the principles of driving. Similarly, those who have to make assessments about a piece of music will make them on the piece as a whole, not on selected parts of it. Chaplen (1970: xxvii) criticized isolated skills tests from this point of view, arguing that "it seems unlikely that measurements of the component Skills most commonly isolated can provide either singly or aggregate, a satisfactory measurement of the gestalt."

This is a view shared by Savignon (1972) who found that grammatical competence was not by itself a good predictor of communicative skills. Kelly (1978) argued that if the goal of applied linguistics were seen as the analysis of meaning, for example, the recognition of the context-specific meaning of an utterance as distinct from its system-giving meaning, then applied linguists should be more interested in the development and measurement of ability to take part in specified communicative performance, the production of and comprehension of coherent discourse, rather than in linguistic competence. This echoed Spolsky's (1968) earlier point that, perhaps, instead of attempting the mastery of grammar and lexis, we would be better employed in testing the person's ability to perform in a specified socio-linguistic setting.

Other testing specialists came out with similar views.

Rea (1978: 51) stated that:

... although we would agree that language is a complex behaviour and that we would generally accept a definition of overall language proficiency as the ability to function in a natural language situation, we still insist on, or let others impose on us, testing measures which assess language as an abstract array of discrete items, to be manipulated only in a mechanistic way. Such tests yield artificial, sterile and irrelevant types of items which have no relationship to the use of language in real life situations.

Morrow (1979) argued that if we were to assess proficiency, (i.e., potential success in the use of the language in some general sense), it would be more valuable to test for a knowledge of and an ability to apply the rules and processes by which these discrete elements are synthesized into an infinite number of grammatical sentences and selected as being appropriate for a particular context, rather than simply to test knowledge of the elements alone.

For Morrow (1979: 145),

...knowledge of the elements of a language in fact counts for nothing unless the user is able to combine them in new and appropriate ways to meet the linguistic demands of the situation in which he wishes to use the language.

II. The Psycho-Sociolinguistic Approach

In response to a feeling that 'discrete-point' tests were insufficient indicators of language proficiency, the testing pendulum on the whole swung in favour of global tests in the 1970s, into what Spolsky (1976) termed the psycholinguistic-sociolinguistic era, an approach to measurement that was in many ways contrary to the allegedly atomistic assumptions of the 'discrete-points' tests (Davies 1978).

Read (1981:61) succinctly described the psycholinguistic-sociolinguistic approach:

From a psycholinguistic perspective, language came to be seen as a well-defined taxonomic structure and more of a dynamic, creative, functional system. It was recognized that natural language contains a considerable amount of redundancy, so that it is difficult to show that any single linguistic unit is indispensable communication... The sociolinguistic contribution centers on the concept of communicative competence, which represents a broadening of Chomsky's notion of competence to cover not only knowledge of rules for forming grammatical sentences but also rules for using those sentences appropriately with different contexts... Thus the psycholinguistic and sociolinguistic perspectives have enlarged the basis on which the validity of a test is to be judged.

Oiler (1979) mentioned that global integrative tests such as dictation and cloze went beyond the measurement of a limited part of language competence achieved by 'discrete point tests towards testing the receptive skills , that such

tests could measure the ability to integrate disparate language skills in ways which more closely approximated the actual process of language use. For Oiler (1979: 37):

The concept of an integrative test was born in contrast with the definition of a discrete point test. If discrete items take language skill apart, integrative tests put it back together . Whereas discrete items attempt to test knowledge of language one bit at a time, integrative tests attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills.

Oiler maintained that linguistic tests such as cloze require 'performance' under real-life constraints, for example, time; they are at least a guide to aptitude and potential for communication, even if they do not test communication itself. They are also practical to administer, economical to set and mark, and have respectable reliability figures associated with them. Although such tests are global in that they require testees to exhibit simultaneous control over different aspects of the language system, they are nevertheless indirect. They might integrate disparate language skills in ways which more closely approximate actual language use, but one would say that their claim to reach communicative validity remains uncertain, as only direct tests which simulate relevant authentic communication tasks can claim to reflect actual communicative interaction. From a communicative point of view, these tests show insufficient attention to the importance of the productive and receptive processing of discourse, arising out of the actual use of language in a social context with all the attendant

performance constraints, for example, the interaction based nature of discourse unpredictability and behavioral outcomes (this can be seen in Morrow 1979 and Moller 1981).

Work by Alderson (1978), however, has raised important questions about the validity of the integrative measures as testing devices. He demonstrated that there is no such thing as "the cloze test", and even in using the same passage, results are affected by altering the point where the deletions are started from, or by using a different nth rate deletion.

According to Davies (1981), there is evidence in the literature that format of a task can unduly affect the performance of some candidates (this can be seen in the work of Murphy 1978, 1980 and Weir 1983). This makes it necessary to include a variety of test formats for assessing each construct rather than rely on a single overall measure such as cloze. In this way, candidates might be given a better chance of demonstrating potentially different abilities.

Both Rea (1978) and Morrow (1979) have emphasized that although indirect measures of language abilities such as close or dictation claim extremely high standards of reliability and concurrent validity as established by statistical techniques, their claim to other types of validity remains uncertain.

Morrow (1979) cited as evidence for this the fact that neither cloze nor dictation offers the opportunity for spontaneous production by the candidate and the language norms which are followed are those of the examiner (or original author of the text), not of the student. Neither testing procedure offers the possibility for oral or noncontrolled written production, and since the oral and written skills are generally held to be highly important, some means of assessing them reliably in communicative situations should be found. Although integrative measures appear to correlate highly with other similar measures of general language

proficiency, there is empirical evidence that doze correlates only moderately with tests of written production (as mentioned by Weir et al. 1978) and with spoken production (Vollmer 1981). Given that the tests concerned are reliable, this would suggest the possibility that proficiency in these areas cannot be adequately predicted by a test of overall proficiency. Morrow also claimed both doze and dictation to be fundamentally suspect since they are tests of underlying ability (competence) rather than actual performance. In other words, they depend basically on knowledge of the language rather than the ability to operate this system in authentic settings.

Concerning the doze test, Carroll (1980:9) reached the same conclusion:

...this (doze test) is still essentially usage based. The task does not represent genuine interactive communication and is, therefore, only an indirect index of potential efficiency in coping with day to day communicative tasks.

Even if it were decided that indirect tests such as doze were valid in some sort of derived fashion, it still remains true that performing on a cloze test is not the same sort of activity as reading. The pedagogical consequences of including this type of test measure in a battery of tests might be harmful in that this would result in candidates being taught specifically to handle indirect assessment tasks in preference to teaching them to cope with more realistic tasks.

Kelly (1978:241) made the observation that some candidates may manage to succeed in the indirect task by a training of a certain kind and thus invalidate the test:

... Indirect tests are subject to attacks on their validity in those cases where it is possible to by-pass the ability

in question and develop proficiency in the assessment task alone.

He (ibid: 245-246) also noted that:

Analysis of a student's responses to an indirect test will not provide any relevant information as to the reasons for the student's difficulties in the authentic task, of which one assumes, the indirect test is a valid and reliable measure. By their very nature, indirect tests can provide evidence for level of achievement, but cannot diagnose specific areas of difficulty in relation to the authentic task.

In this sense, integrative tests (such as cloze) only tell us about a candidate's linguistic competence. They do not tell us anything directly about a student's performance ability, and their main value in their unmodified form is in designating competence levels rather than relating candidates' performance to any external criteria. They are only of limited use where the interest is in what the student can or cannot do in terms of the various language tasks they may face in real life situations. High correlations between cloze and other measures may only reflect that they are measuring different skills which are highly correlated among individuals; however, this does not mean that there will be no individuals whose performances in the various skills differ considerably.

The deficiencies in the type of information the 'discrete point' approaches of the psychometric- structuralist era and the more integrative approaches of the psycholinguistic-sociolinguistic era can provide bring about a need to

investigate the 'communicative paradigm' to see whether this approach might prove more satisfactory.

III-The Communicative Approach

The communicative language testing approach. It considers language to be interactive, purposive, authentic, contextualized, and based **and** assessed in terms of behavioral outcomes. This approach emphasizes evaluation of language use rather than usage. It is based on the premise that language is first and foremost a tool for communication. From this perspective, tests designed to assess student proficiency can be tailored to include items which possibly measure the students' communicative ability in all levels of language.

1-Communicative language Ability

We would like to compare foreign language ability to the ability to drive a car. In both cases, it seems to us that there are a number of enabling micro-skills which have to be learnt in isolation or perhaps with a bit more attention paid to each one at a time so that one can satisfactorily achieve the intended goals. In the case of driving, these micro-skills are starting the engine, using the breaks, changing gear, signaling intention to turn right or **left...In** the case of using a foreign language, micro-skills include relating different parts of a text, summarizing, expressing different feelings through intonation, changing the topic of a conversation...According to Widdowson (1983: 30), these skills are "the backstage facilitators" which have to be mastered in order to be disregarded while "the play is in action" —to use yet another analogy.

This notion of the learning process brings us to the problem of assessment. It seems only fair that a teacher or a course planner would like to

know the extent to which each learner is capable of making use of micro skills in order to achieve a certain goal. It might be claimed that the only way to do this is to follow the learner around in order to observe his/her actual behaviour and assess this performance. In the case of driving, this "realistic" assessment would mean sitting in a car and letting the learner drive around, coping with all sorts of problems. Not only in relation to driving, but in terms of language learning as well, such testing method presents a series of problems, as pointed out by Skehan (1988). To start with, it is impractical, biased, unethical and unreliable. Besides, how can one decide that a representative sample of behaviour has been observed, so that a judgment can be made? Thirdly, in the case of language instruction, (and unlike driving) we are very often concerned with progress in relation to a set of objectives. Performance, then, is assessed in terms of different stages of the learning process (i.e. the process of integrating different parts of the whole with different levels of mastery), and not against a "competent users' model of language proficiency.

What follows then is that when designing and implementing language tests, we have to work according to a set of underlying principles, the same sort of framework which guides the setting of pedagogic objectives and methodological procedures. In other words, as stated by Skehan (1988:211)

We can only decide what and how test by recourse to theory- theory which guides and predicts how an underlying communicative competence is manifested in actual performance; how situations are related to one another, how competence can be assessed by examples of performance on actual tests; what components communicative competence has; and how these inter-relate.

According to Candlin (1986:40)

Communicative language ability (CLA) can be described as consisting of both knowledge, or competence, and the capacity for implementing, or executing that competence in appropriate contextualized communicative language use. ...the ability to create meanings by exploring the potential inherent in any language for continual modification in response to change, negotiating the value of convention rather than conforming to established principle. In sum, ...a coming together of organized knowledge structure with a set of procedures for adapting this knowledge to solve new problems of communication that do not have ready-made and tailored solutions.

The framework of Communicative Language Ability (CLA) proposed by Candlin includes three components: language competence, strategic competence, and psycho physiological mechanisms. Language competence comprises, essentially, a set of specific knowledge components that are utilized in communication via language. Strategic competence is the term used to characterize the mental capacity for implementing the components of language competence in contextualized communicative language use. Strategic competence thus provides the means for relating language competencies to features of the context of situation in which language use takes place and to the language user's knowledge structures (sociocultural knowledge, 'real world' knowledge). Psycho physiological mechanisms refer to the neurological and psychological processes involved in the actual execution of language as a physical phenomenon (sound, light). The interactions of these components of

CLA with the language use context and language user's knowledge structures are illustrated in the Figure below.

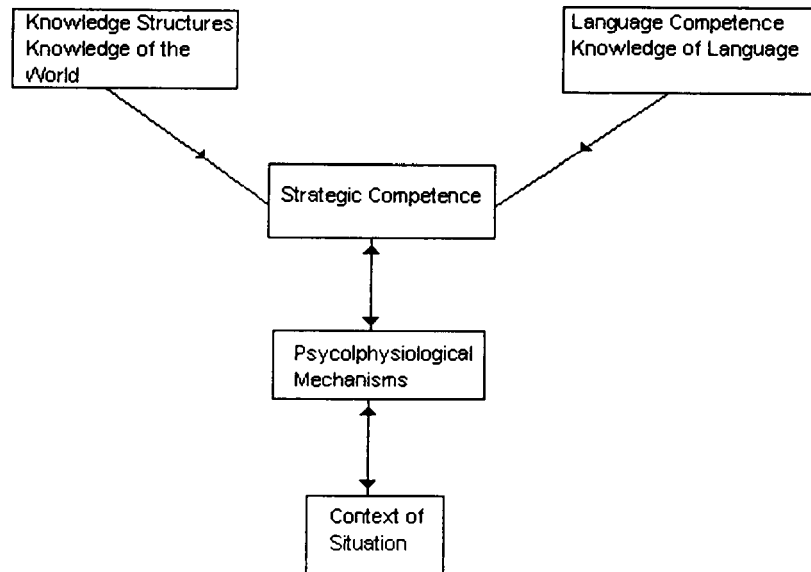


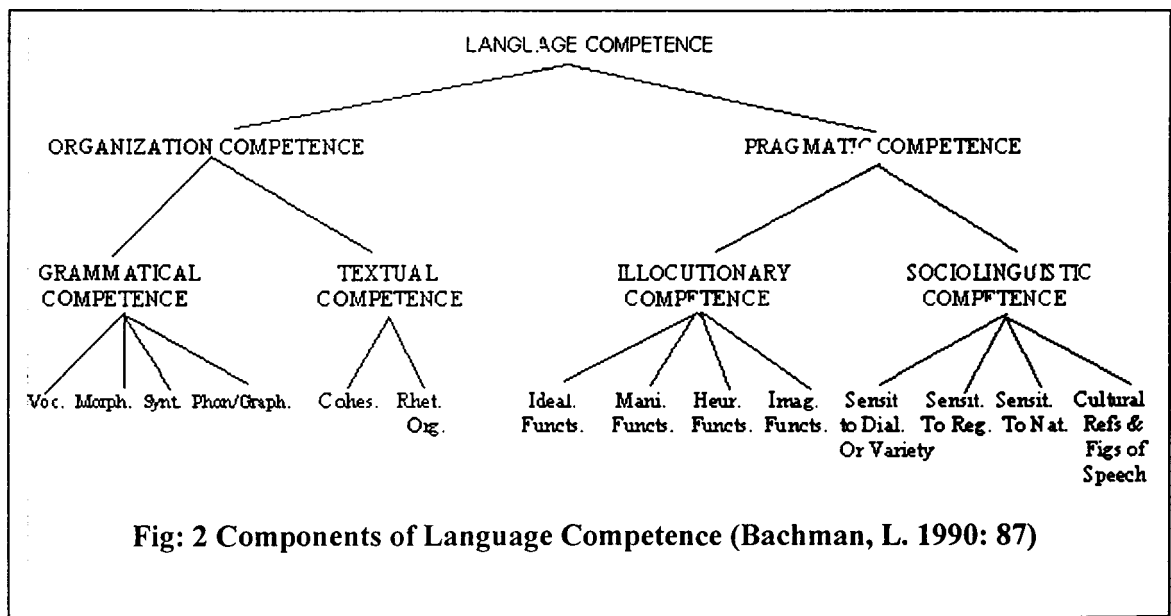
Figure: 1 Components of Communicative Language Ability in Communicative Language Use.

The idea of competence, as in Canale and Swain framework (Canale and Swain: 1980, extended in Canale: 1983) relates to underlying knowledge of different aspects of language. They define competence in terms of four areas of knowledge and skill:

- **grammatical competence**, which has to do with mastery of the language code —lexis, syntax and phonology;
- **sociolinguistic competence**, which refers to the appropriacy of language in relation to different sociolinguistic contexts;
- **discourse competence**, which relates to the organization of discourse in terms of cohesion and coherence; and
- **strategic competence**, which encompasses the verbal and non verbal strategies used to compensate for breakdowns in communication or to enhance the effectiveness of communication. It is the part which performs

assessment, planning and execution functions in determining the most effective means of achieving a communicative goal.

Bachman's (1990) model of communicative competence incorporates the ideas discussed above and adds other different components. He uses the term communicative language ability as a general term, to include language competence, strategic competence and psycho physiological mechanisms and language competence includes what Canale and Swain, described as grammatical, discourse and sociolinguistic competence, and Bachman explains his own model of language competence as demonstrated in the diagram below.



The description of language competence presented here builds upon empirical findings by grouping morphology, syntax, vocabulary, cohesion, and Organization under one component, organizational competence, Pragmatic

competence, but also those abilities related to the functions that are performed through language use. Language competence is classified into two types: organizational competence and pragmatic competence. Each of these, in turn, consists of several categories. The components of language competence are illustrated model, and as with any metaphor, it captures certain features at the expense of others. This diagram represents the hierarchical relationships among the components of language competence, at the expense of making them appear as if they are separate and independent of each other. However, in language use, these components all interact with each other and with features of the language use situation. Indeed, it is this very interaction between the various competencies and the language use context that characterizes communicative language use.

2-Characteristics of a Communicative Language Test

Brown (1987:230) elaborates on the characteristics of a communicative language test:

A communicative test has to meet some rather stringent criteria. It has to test for grammatical, discourse, sociolinguistic, and illocutionary competence as well as strategic competence.

Such a test has to be pragmatic in that it requires the learner to use language naturally for genuine communication and to relate to thoughts and feelings, in short, to put authentic language to use within a context. It should be direct (as opposed to indirect tests which may lose validity as they lose content validity). It should test the learner in a variety of language functions.

An important observation in the previous quotation is that in testing communicative performance, test items should measure how well students are able to engage in meaningful, purposeful, and authentic communicative tasks. Students must have a good performance linguistically and communicatively. That is, they must have a good command of the components involved in communication. The best exams in this communicative era, Madson (1983) comments, are those that combine the various sub skills necessary for the exchange of oral and written ideas. He asserts that communicative tests need to measure more than isolated language skills, to comprehensively indicate how well a person can function in another language.

The communicative approach to language testing is sometimes linked to the integrative approach in that it involves the testing of language in context, and is thus concerned primarily with meaning and the total communicative effect of discourse. However, although both approaches emphasise the importance of the meaning of utterances rather than their form and structure, there are nevertheless fundamental differences between the two approaches. We can notice that communicative tests are concerned primarily (if not totally) with how language is used in communication. Consequently, their major aim is to incorporate tasks which approximate as closely as possible to those facing the students in real life. Success is judged in terms of the effectiveness of the communication which takes place rather than formal linguistic accuracy. Language use is often emphasised to the exclusion of language usage as it is concerned with how people actually use language for a multitude of different purposes while usage' concerns the formal patterns of the language.

In practice, however, some tests of a communicative nature include the testing of usage and assess the ability to handle the formal patterns of the target

language. Some supporters of the communicative approach would argue that communicative competence can never be achieved without a considerable mastery of the grammar of a language.

The attempt to measure different language skills in a communicative test is based on a view of language referred to as the "divisibility hypothesis". This hypothesis holds that Communicative testing results in an attempt to obtain different profiles of a learner's performance in the language: the learner may, for example, have a poor ability in using the spoken language in informal conversations, but may score quite highly on tests of reading comprehension. In this sense, communicative testing draws heavily on the recent work on aptitude testing where it has long been claimed that the most successful tests are those which measure separately such relevant skills as the ability to translate news reports, the ability to understand radio broadcasts, or the ability to interpret speech utterances. The score obtained on a communicative test will thus result in several measures of proficiency rather than simply one.

To have a good understanding of what is meant by communicative testing, it is necessary to examine more closely the concept of communicative competence. For Hymes (1972), communicative competence is a two-dimensional model, comprising a 'linguistic' element—the ability to use the language and a 'sociolinguistic' element—the knowledge which underlay actual performance. Most subsequent models have included considerations of a sociolinguistic dimension, but relatively little is in fact done and known about the communicative paradigm, and adequately developed theories of communicative language use are not yet available. This is not to say that we must wait for completion of such theories before appropriate procedures can be developed. Rather, we need to investigate systematically some of the available

hypotheses about language use and try to operationalise these for testing purposes.

Morrow (1979) and Canale et al. (1980) provided a useful starting point for a clarification of the terminology necessary for forming a more definite picture of the construct of communicative testing. They argued that communicative language testing as well as being concerned with what the learner knows about the form of the language and about how to use it appropriately in contexts of use (competence), must also deal with the extent to which the learner is actually able to demonstrate this knowledge in a meaningful communicative situation (performance), i.e., what he can do with the language, or as Rea (1978:4) put it, "...his ability to communicate with ease and effect in specified sociolinguistic settings." . They felt that a distinction needed to be made between communicative competence and communicative performance.

They explain that communicative competence includes grammatical competence (knowledge of the rules of use and rules of discourse) and strategic competence (knowledge of verbal and non-verbal communication strategies). The model was subsequently updated by Canale (1983), who proposed a four-dimensional model comprising linguistic, sociolinguistic, discoursal and strategic competences. Concerning communicative performance, they state that the performance tasks candidates are faced with in communicative tests should be representative of the type of task they might encounter in their own real-life situation and should correspond to normal language use where an integration of communicative skills is required, with little time to reflect on, or monitor language input and output. The criteria employed in the assessment of communicative performance on these tasks should relate closely to the effective communication of ideas in that context. Models such as these provide a useful

framework for the design of language tests, but it must be emphasized that they are still in need of validation.

Like the separate testing of the skills in the structuralist approach, it is felt in communicative testing that sometimes the assessment of language skills in isolation may have only a very limited relevance to real life. For example, reading would rarely be undertaken solely for its own sake in academic study, but rather for subsequent transfer of information obtained to writing or speaking. Since language is decontextualised in a psychometric-structural test, it is often a simple matter for the same test to be used globally for any country in the world. Communicative tests, on the other hand, must of necessity reflect the culture of a particular country because of their emphasis on context and the use of authentic materials. Not only should test content be totally relevant for a particular group of testees, but the tasks set should relate to real-life situations, usually specific to a particular country or culture.

We can notice that there has been a remarkable shift in emphasis from the linguistic to the communicative dimension. The emphasis is no longer on linguistic accuracy, but on the ability to function effectively through language in particular contexts of situation. Cooper's (1968) view that existing test frameworks, because they concentrated on linguistic competence, might fail to assess a person's communicative ability, was taken up by Morrow (1979:149) who argued that traditional tests did not give

...any convincing proof of the candidate's ability to actually use the language, to translate the competence (or lack of it) which he is demonstrating into actual performance 'in ordinary situations' i.e. actually using

the language to read, write, speak or listen in ways and contexts which correspond to real life.

Carroll (1980:7) expressed a similar view:

...the prime need of most learners is not for a theoretical or analytical knowledge of the target language, but for an ability to understand and be understood in that language within the context and constraints of particular language-using circumstances.

For him (op. Cit., 7)

...the ultimate criterion of language mastery is therefore the learner's effectiveness in communication for the settings he finds himself in.

We can say that communicative testing is one which samples behaviour in a single setting with no intention of generalizing beyond that setting- any other test is bound to concern itself with competence. The very act of generalizing beyond the setting actually tested implies some statements about abilities to use the language and/or knowledge of it. It would be then more accurate in discussing communicative language testing not to claim to be doing anything more than evaluating samples of performance, in certain specific contexts of use, created under particular test constraints, for what they can tell us about a candidate's underlying communicative competence.

For Kelly (1978:350)

To take part in a communicative event is to produce and/or comprehend discourse in the context of situation and under the performance conditions that obtain. It is the purpose of a proficiency test to assess whether or not candidates are indeed capable of participating in typical communication events from the specified communication situation (s).

It is, however, difficult to see how competence (knowing about using a language) might be evaluated except through its realization in performance. Only performance can be directly observed and hence evaluated. All linguistic behaviour, even completing multiple choice tests of phoneme discrimination, necessarily involves performance. In practice, a clear distinction between performance and competence will be difficult to maintain.

Rea (1985) argues that all tests are more satisfactorily seen as tests of performance that are to varying degrees communicative or non-communicative, or (in Widdowson's dichotomy) use or usage based. Rea also distinguishes between items as meaning dependent or independent, and describes how the former can be subdivided according to whether they involve a context determined response or not. Considering Rea's point of view, we might accept that communicative performance relates to the transmission and reception of particular meanings in particular contexts, and what can be tested is the quality and effectiveness of the performance observed in these circumstances.

In the current testing literature, there is a strong emphasis on the importance of test purpose, and it is held that no one solution can accommodate

the wide variety of test scenarios. It is argued that differentiated tests in different skills areas need to be made available for evaluating different groups of testees with different target situation needs. To measure language proficiency adequately in each situation, account must now be taken of where, when, how, with whom, and why the language is to be used, and on what topics, with what effect.

One can imagine these characteristics being relatively easily attended to in performance —based proficiency tests geared towards the assessment of ability to perform a restricted set of very well-defined tasks. These are, for example, the sort of tests described by Wesche (1987), which rather than trying to establish overall proficiency levels, set out to judge the extent to which test-takers are ready to carry out certain specific activities. Even then, however, aspects such as unpredictability, response to context, purpose of interaction and authenticity of materials will be very much determined and restricted by not only the range of abilities being tested but also by the limitations of length and time which are imposed on a test.

As for the progress and achievement tests being focused on in this thesis, it seems to us that most of the characteristics outlined by Morrow can be attended to by sampling a syllabus which claims to be designed to foster the development of communicative language ability. (Alderson, 1983) However, sampling such a syllabus —if it exists- is not enough as a guarantee that any test will be an effective measure of language ability.

We believe that the most authentic behaviour in a language test is to display knowledge which can then be assessed by an examiner. Nevertheless, if tests are going to mirror the teaching syllabus and classroom activities, it seems that the texts used will be of the type learners are familiar with. As for question

types, some distinction will perhaps be made between those which are more suitable for teaching and those which are probably better testing instruments. A cloze exercise, for example, seems to be the kind of technique which goes with testing more than with teaching.

In relation to assessment of outcome, it seems that one must accept the premise that linguistic competence (in Bachman's terms) or communicative competence (as defined by Canale and Swain) is divisible and subject to assessment of their different components. Hence, tasks which can be perceived as global and self-contained as pieces of discourse, can lend to interpretations of focus on different aspects of language ability. This is, in fact, the approach taken by Milanovic (1988) in the design of large-scale achievement tests for adult beginners. Each one of these tests deals with sub-skills (based on Munby: 1978 and Carroll: 1978), and different aspects of competence, defined according to the different levels of the course. In the case of writing and speaking, production is evaluated holistically on the basis of achievement of goal, as well as with reference to different aspects of competence (e.g. linguistic, discoursal, etc.)

The reading and listening tests try to assess comprehension with a minimum amount of production. The Use of English test focuses not only on aspects of lexis and syntax but on sociolinguistic appropriacy as well.

The questions which the test designers have to face, as mentioned by Rea (1985) are:

- Is it really possible to maintain the theoretical distinction between language competence and strategic competence as in the Bachman model?
- If what one can observe through a test is actual performance, how is one to assess this performance? Simply on the basis of achievement of communicative goal? But, then, is communicative goal an absolute

concept, or is it open to interpretation in relation to the speakers involved and the context of situation?

- If performance is judged not to be satisfactory, how does one know whether the drawback was in the area of linguistic competence or whether the student simply could not access some existing knowledge?
- What kind of test will best prompt students' performance so that they can use underlying competence?

Morrow (1979) has listed a number of characteristics of language use which, according to him, are not taken into account by conventional tests (which he does not define very clearly, apart from mentioning some testing techniques such as cloze and dictation) :

- interaction-based, i.e., presupposing more than one participant ;
- unpredictability, in terms of the directions communication can take ;
- context, both linguistic and of the social situation ;
- purpose ;
- performance, i.e., constraints such as time pressure, memory failure or emotional states which can hinder performance ;
- authenticity, i.e., avoidance of specially created materials ; and,
- behaviour-based, in the sense that successful outcome is judged on the basis of communicative goal.

As a concluding statement about the three major testing approaches, we would say that the important role of context as a determinant of performance ability is stressed and an integrative approach to assessment as against a decontextualised approach is advocated. For Oller (1973, 1979), the higher the level at which language is contextualized, the more effective language perception, processing and acquisition are likely to be. The variability in

performance according to the discourse domain or type of task involved is recognized with the attendant implications this might have for test length. Authenticity of both texts and test tasks is regarded as something worth attempting to pursue despite the problems involved. The fact that communicative performance is affected by prior knowledge/experience/abilities is accepted along with the implications of this for test specificity. The possible influence of test method on trait estimation is increasingly recognized, if not yet fully understood. The principle of eliciting the student's best performance by minimizing the effect of the measurement technique on this is advocated. The net result of these considerations is that different tests are constructed to match different purposes and these instruments are no longer uniform in content or method.

Conclusion

The different approaches to language testing, the structuralist, the psychosociolinguistic and communicative) are rather complementary, in the sense that each approach responds to and serves specific needs of a particular population of learners during a period of time. They all have their limitations in terms of criteria of validity, reliability and practicality. The useful approach would be an eclectic one, which incorporates positive features from other approaches.

CHAPTER FOUR

TESTING ORAL PROFICIENCY

Introduction	99
I-The Oral Proficiency	100
II-Oral Proficiency Testing	103
1-Approaches to Testing the Oral Proficiency	103
2- Specificity of Oral Proficiency Testing	106
3-Criteria of Testing Oral Proficiency	107
4-Improving the Reliability of Oral Examinations	111
5- The Assessment Scales	118
III -Procedures of Testing Oral Proficiency in the Department of English	127
Conclusion	129

Introduction

The traditional view that written skills have greater importance than oral skills still prevails. In the last century, the arguments for a strong emphasis on oral skills in foreign language teaching were not very strong. Few people travelled, business was not as international as it is today, communication with other countries was mainly by mail, and for many students the main attraction of studying a foreign language was probably the opportunity to read literature in the target language. Teaching methods were chosen accordingly: there was an emphasis on writing, especially the translation of literary texts, and a minor role for oral skills.

In the 1990s the situation has changed radically. Travel abroad is now within the reach of almost everybody, business is increasingly international, and Communication between countries is more in the form of direct contact by telephone. An interest in literature is also much less common among students of foreign languages, who include among their numbers those combinations with Law, Business Studies, Linguistics, Social Sciences, etc. Many students are quite likely to be more familiar with foreign language films, television series and pop-stars than with foreign literature. It is therefore no surprise that a greater number of educational institutions nowadays have started to stress the importance of oral skills. Schools and many other educational institutions have responded to the perceived changes in the needs of language learners mainly by teaching foreign languages to a far larger number of learners than previously, and by changing methods, moving away from the traditional translation and grammar approach towards 'communicative', mainly oral, skills. The focus of debates over oral assessment has taken a great importance in the field of teaching English.

Against this background, developments are taking place in our higher system of education, with the introduction of new courses and combination involving foreign languages, and changes to the content of existing courses (adopting the LMD system). In this study we would like to find out and look at how the growing importance of the oral proficiency in foreign languages is reflected in the testing, and to some extent also in the teaching, of oral skills for first year university students.

I-The Oral Proficiency

Language researchers openly acknowledge the lack of consensus about what the oral proficiency encompasses. Cummins (1984), for example, states that the nature of oral proficiency has been understood by some researchers as consisting of many separate language components and by others as consisting of only one global factor. Valdés and Figueroa (1994:34) indicate that:

...what it means to know a language goes beyond simplistic views of good pronunciation, "correct" grammar, and even mastery of rules of politeness. Knowing a language and knowing how to use a language involves a mastery and control of a large number of interdependent components and elements that interact with one another and that are affected by the nature of the situation in which communication takes place.

Oiler and Damico (1991) succinctly state that the nature and specification of the elements of oral proficiency have not been determined and there continues to be debate among academicians and practitioners about the definition.

The complexity of language and the lack of consensus as to the exact nature of oral proficiency is critical for one fundamental reason: each language proficiency test should be based on a defensible model or definition of oral language proficiency. The question is which definition to use? Oral proficiency tests have been developed on the basis of a plethora of definitions and theories. Additionally, the test developer may indicate that a test is based on a particular model of oral language proficiency but, it remains to be seen just how successfully the model was actually operationalized in the form of a test. In other words, describing the theoretical model of oral language proficiency in a technical manual does not mean that the test exemplifies the model.

We can say that a fully English proficient student is someone who is able to use English to ask questions, to understand teachers, to read materials, to test ideas, and to challenge what is being asked in the classroom. He /She is also the one who has the ability to use oral language appropriately and effectively in learning activities (such as peer tutoring, collaborative learning activities, and question/answer sessions) within the classroom and in social interactions within the school. Canale (1994) offers an equally practical definition of English oral proficiency. Her definition of oral proficiency is predicated on a socio-theoretical foundation. What this means is that language is more than just the sum of discrete parts (for example, pronunciation, vocabulary, grammar). It develops within a culture for the purpose of conveying the beliefs and customs of that culture. Anyone who has ever tried to translate an idiom from one language to another understands this premise. A "bump on a log" in English means someone who is lazy or a does-nothing, but the non-English speaker has to assimilate the idiom rather than the meaning of each individual word in order to make sense of the phrase. In this sense Canale (ibid: 60) says that oral proficiency is:

- dynamic and contextually-based (varies depending upon the situation, status of the speakers, and the topic),
- discursive (requires connected speech),
- and requires the use of integrative skills to achieve communicative competence.

In other words, oral proficiency is a coherent orchestration of discrete elements, such as vocabulary, discourse structure and gestures to communicate meaning in a specific context.

If we consider the kinds of linguistic abilities that underlie the successful academic performance of students, we see that students must be able to orally respond to teacher and peer queries for information, ask probing questions, and synthesize reading material. They must be able to understand routine aural instructions in a large group setting and peer comments in a small group setting. Moreover, the successful language user also knows the social and cultural rules governing these and many other language mediated activities.

Valdés and Figueroa (1994:62) maintain that oral proficiency testing should require this kind of contextualized language processing. They take the position that it is feasible to:

... identify the levels of demand made by such contexts and the types of language ability typical of native, monolingual English speaking learners who generally succeed in such contexts. From these observations, one could derive a set of criteria against which to measure the abilities of non-native English speaking children in order to decide whether to educate them in English or their home language.

The reason for this recommendation is obvious. An oral English language proficiency test score is intended to assist educators in making an accurate judgment regarding which learners need English language instructional assistance or no longer need such assistance. Making such a judgment becomes difficult when the language tasks underlying the test score bear little resemblance to the language tasks characteristic of a mainstream classroom.

II-Oral Proficiency Testing

1-Approaches to Testing Oral Proficiency

Oiler and Damico (1991:82) indicate that oral proficiency tests can be associated with three schools of thought: Discrete-point testing, Integrative Testing and Pragmatic Testing.

Discrete point is based on the assumption that language proficiency consists of:

... of separable components of phonology, morphology, lexicon, syntax, and so on, each of which could be further divided into distinct inventories of elements (e.g., sounds, classes of sounds or phonemes, syllables, morphemes, words, idioms, phrase structures, etc.)

Following the discrete point model, an oral proficiency test typically uses testing formats such as phoneme discrimination tasks where the test taker is required to determine whether or not two words presented aurally are the same or different (for example, /ten/ versus /den/). A similar example might be a test designed to measure vocabulary which requires the test taker to select the appropriate option from a set of fixed choices. The authors conclude that the

weaknesses leading to the demise of such thinking centred upon evidence such as:

- the difficulty of limiting oral testing to a single skill for example, Multiple Choice Questions (MCQ)
- the difficulty of limiting oral testing to a single linguistic element (for example, vocabulary) without involving other domains (for example, phonology);
- and the difficulty of measuring oral proficiency in the absence of any social context or link to human experience.

The Integrative approach of testing requires that oral language proficiency be assessed "in a fairly rich context of discourse".(Damico and Oller opcit 1991: 83) This assumption was based on the belief that oral language processing or use entails the simultaneous engagement of more than one language component (for example, vocabulary, grammar, gesture) and skill (for example, listening, speaking). Following this logic, an integrative task might require the test-taker to listen to a story and then retell the story or to read the story and then summarize the story.

Pragmatic language testing differs from Integrative testing in one fundamental way: an ostensible effort is made to link the language testing situation with the test-taker's experience. As Oller and Damico (1991) state, normal language use is connected to people, places, events and relations that implicate the whole continuum of experience and is always constrained by time or temporal factors. Consequently, pragmatic oral language tasks are intended to be as "real life" or authentic as possible. That might require the test-taker to engage in a listening task like an integrative task, but under the contextual and temporal conditions that generally characterize this activity. From a pragmatic perspective, language learners do not generally listen to audio-taped stories; they

more commonly listen to adults or competent readers read stories. In this sense a story-retell listening task which uses a tape-mediated story falls short of meeting pragmatic criteria. A pragmatic approach to story retelling might take on the following features:

- normal visual input is provided (for example, the reader's gestures, the print on the page, an authentic number of story linked pictures in the text);
- time is managed differently in that the learner may have opportunities to ask questions, make inferences, or react in a normal way towards the content of the story; and
- the story, its theme, the reader, and the purpose of the activity form part of the learner's experience.

Oiler and Damico (1991:85) make an interesting observation regarding the power of pragmatic oral proficiency testing. The researchers explain that, what is more important about pragmatic tests, and what is yet to be appreciated fully by theoreticians and practitioners is that all the goals of discrete point items (for example, diagnosis, focus, isolation) are better achieved in the full rich context of one or more pragmatic tests. As a method of linguistic analysis, the discrete point approach has some validity, but as a practical method for assessing language abilities, it is misguided, counterproductive, and logically impossible. In other words, if the intent is to measure the learner's proficiency in the areas of grammar, vocabulary and pronunciation, for example, this is best achieved through a pragmatic language approach which seems to meet the demands of educators. Nonetheless, educators will be limited to the use of currently available tests which may or may not measure oral proficiency in a pragmatic, "real life" manner. All in all, we can say that opting for one approach or another will depend on the situation and the objectives to be attained

2- Specificity of Oral Proficiency Testing

Many teachers feel comfortable setting pencil-and-paper tests. Years of experience marking written work have made them familiar with the level of written competence pupils need in order to succeed in a specific standard. However, they often feel much less secure when dealing with tests which measure speaking and listening even though these skills are regarded as essential components of a diagnostic test which measures overall linguistic proficiency. Although some of the second-language English learners often come from an oral rather than a written culture, and so are likely to be more proficient in this mode of communication, at least in their own language, speaking in English may be a different matter. The lack of proficient, well trained teachers and an authentic environment may impede students' acquisition of oral proficiency. Therefore, identifying and providing the correct learning environment to the student is all the more challenging and important.

How can one succeed to set a test which does not intimidate but encourages learners to provide an accurate picture of their oral ability. In replying to this question, one needs to consider briefly the findings of researchers working in the field of language testing. According to Madsen (1983:147) "The testing of speaking is widely regarded as the most challenging of all language tests to prepare, administer and score," This is especially true when examining University first year students who have just started to apply for a degree in English.

Theorists suggest three reasons why the oral proficiency test is so different from more conventional types of tests. Firstly, the nature of the speaking skill itself is difficult to define. Because of this, it is not always easy to establish criteria to evaluate a speaking test. Is "fluency" more important than

"accuracy," for example? If we agree that fluency is more important, then how will we define this concept? Are we going to use "amount of information conveyed per minute" or "quickness of response" as our definition of fluency? Second, when testing beginning-level speakers of English, which involves getting them to speak in the first place, and then defining the role the tester will play while the speaking is taking place, it is important to develop relevant elicitation procedures which will prompt speakers to demonstrate their optimum oral performance. The tester will therefore need to act as a partner in the production process, while at the same time evaluating a number of aspects about this production. A third set of difficulties emerges if one tries to treat an oral test like any other more conventional test. Madsen (1983:159) argues

In the latter, the test is often seen as an object with an identity and purpose of its own, and the learners taking the test are often reduced to subjects whose only role is to react to the test instrument.

In oral tests, however, the priority is reversed. The people involved are important, not the test, and what goes on between tester and testee may have an existence independent of the test instrument.

3-Criteria of Testing Oral Proficiency

Criteria of testing oral proficiency have received from different educational institutions increasing attention since the 1970s.

- Assessment of oral communication should view competence in oral proficiency as a gestalt of several interacting dimensions. At a minimum, all assessments of oral communication should include an assessment of knowledge (understanding communication process, comprehension of the elements, rules,

and dynamics of a communication event, awareness of what is appropriate in a communication situation), an assessment of skills (the possession of a repertoire of skills and the actual performance of skills), and an evaluation of the individual's attitude toward communication (for example, value placed on oral communication, apprehension, reticence, willingness to communicate, readiness to communicate).

- Because oral communication is an interactive and social process, assessment should consider the judgment of a trained assessor as well as the impressions of others involved in the communication act (audience, interviewer, other group members, conversant), and may include the self-report of the individual being assessed.

- Assessment of oral communication should clearly distinguish speaking and listening from reading and writing. A major portion of the assessment of oral communication should require speaking and listening. Directions from the assessor and responses by the individual being assessed should be in the oral/aural mode.

- Assessment of oral communication for all students should include assessment of both verbal and nonverbal aspects of communication and should consider competence in more than one communication setting. A minimum assessment should occur in the one-to-many setting (for example, public speaking, practical small group discussion) and in the one-to-one setting (for example, interviews, interpersonal relations)

- The method of assessment should be consistent with the dimension of oral communication being assessed. While knowledge and attitude may be assessed in part through paper and pencil instruments, speaking and listening skills must be assessed through actual performance in social settings (speaking

before an audience, undergoing an interview, participating in a group discussion, etc.) appropriate to the skill(s) being assessed.

- Instruments for assessing oral communication should describe degrees of competence. (Either/or descriptions such as "competent" or "incompetent" should be avoided) and should attempt to diagnose reasons why individuals demonstrate or fail to demonstrate particular degrees of competence.

- Instruments for assessing each dimension of oral communication competence should clearly identify the range of responses which constitute various degrees of competence. Examples of such responses should be provided as anchors.

- Assessment instruments should have an acceptable level of reliability: test/retest reliability, split-half reliability, alternative forms reliability, inter-rater reliability, and internal consistency.

- Assessment instruments should have appropriate validity: content validity, predictive validity and concurrent validity.

- Assessment instruments should be suitable for the developmental level of the individual being assessed.

- Assessment instruments should be standardized and detailed enough so that individual responses will not be affected by an administrator's skill in administering the procedures.

- Use of competence assessment as a basis for procedural decisions concerning an individual should, when feasible, be based on multiple sources of information, including especially:

- +/- direct evidence of actual communication performance in class and/or other contexts.

- r results of formal competence assessment, and

r measures of individual attitudes toward communication (for example, value placed on oral communication, apprehension, reticence, willingness to communicate, and readiness to communicate).

- Individuals administering assessment procedures for oral communication should have received sufficient training by speech communication professionals to make their assessment reliable. Scoring of some standardized assessment instruments in speaking and listening may require specialized training in oral communication on the part of the assessor.

The results of student oral communication competency assessment should be used in an ethical, non-discriminatory manner for such purposes as:

- + Diagnosing student strengths and weaknesses,
- + Planning instructional strategies to address student strengths and Weaknesses,
- 4- Certification of student readiness for entry into and exit from programs and institutions;
- + Evaluating and describing overall student achievement,
- r Screening students for programs designed for special populations,
- r Counselling students for academic and career options, and
- 4- Evaluating the effectiveness of instructional programs.

No single assessment instrument is likely to support all these purposes. Moreover, instruments appropriate to various or multiple purposes typically vary in length, breadth/depth of content, technical rigor, and format.

4-Improving the Reliability of Oral Examinations

Oral examination have received much criticism in the literature on account of their reliability (the consistency with which the same performance received the same marks from different examiners or on different occasions) and *validity*. Very often, some students show their dissatisfaction about the validity of oral examinations, (the degree to which an examination measures what it sets out to measure) but rather greater dissatisfaction with their reliability.

The reliability of examinations does not, however, depend solely on whether they are written or oral, but rather on how objective they are (an objective test being one in which there is a single correct answer to each question, for example, multiple choice or cloze tests). There are objective oral and written examinations, but most of the questions set to students in higher education, while being highly valid, do not belong to the objective type of test and thus have relatively low reliability. This applies both to essay-type questions (which, at least in language courses, make up the great majority of written examinations) as well as to oral examinations, but it is clearly perceived as a problem particularly in the latter.

Some of the factors already identified as contributing to the unreliability of oral examinations can thus be eliminated, and the more research done on this topic to identify further factors, the more reliable the marking of oral examinations will become. We must, however, admit that oral examinations of the type which we would wish to give in higher education (first year University students, Department of English) will never be among the most reliable tests. Like essay-type questions, oral examinations compensate for their weaknesses in reliability by high levels of validity. This does not mean that we can be complacent about the reliability of either oral or essay-type written examinations: improvements can certainly be made in both.

In order to improve the reliability of oral proficiency tests, we have to pay special attention to test construction, the markers, examination format and the marking scheme.

When devising any test, it is true that certain norms and standards have to be observed. Carroll (1980:13) suggested three phases:

- | | |
|----------------------|-------------------------------------|
| Phase 1: Design | 1. Description of participant(s). |
| | 2. Analysis of communicative needs. |
| | 3. Specification of test content. |
| Phase 2: Development | 4. Realization of test. |
| | 5. Trial application. |
| | 6. Validation and test analysis. |
| Phase 3: Operation | 7. Full-scale application. |
| | 8. Operational use. |
| | 9. Revision of test system. |

Research has shown that trained, experienced examiners can achieve consistent results (a very high correlation coefficient of 0,91 for senior examiners in O-level Italian oral examinations is reported in Francis 1981:18). It is highly doubtful whether such a high degree of correlation could be achieved without any training or coordination of examiners. It is also generally acknowledged that the reliability of marks is considerably enhanced when more than one marker is involved. An ideal number is two or three; larger numbers do not systematically increase reliability (the 'numbers paradox', Underhill 1987:89-90). Reality shows that a majority of oral assessment, in the Department of English regularly

have one examiner present at oral examination. On the other hand, there is a significant number of cases in which the external examiner examines on his/ her own. We strongly believe that this practice needs reconsidering. As a general rule, it is said that the more advanced the level, the longer the oral examination should be; as more advanced learners need longer to demonstrate their greater proficiency, range of skills. However, this presents problems obtaining a satisfactory basis of assessment where more than one examiner is involved and the number of students is large, and from the cost in terms of staff time, fluctuations caused by tiredness which are likely to reduce reliability. We think that the possible solution to this dilemma lies in self- assessment, which is the next point we would like to discuss and propose for our colleagues in the Department of English.

In practice, most examinations consist of a series of tasks (for example, reading aloud, answering questions, discussing a topic, conversation, role play, problem solving), which ideally should include both highly reliable and highly valid parts (cf. Underhill 1987:6). For this purpose, Murphy suggested examples of questions which might be asked of the activities candidates are expected to perform:

- Are the activities set at an appropriate level of difficulty?
- Will the test discriminate adequately between the performances of candidates at different levels of attainment?
- Does the test assess the full range of appropriate skills and abilities, as defined by the objectives of the examination?
- Are the activities unambiguous, giving a clear indication of what the examiner is asking, so that no candidate may take the task to mean something different?

- Is there an excessive overlap in enabling skills or communicative activities being assessed in different parts of the test?
- Can the activities be satisfactorily answered in the time allowed?

In addition, the examiner or the panel might consider the format and the layout of the questions. This is important because a badly laid out question could be the cause of considerable problems for both candidates and examiners. Instructions to candidates need to be as clear and as concise as possible.

Murphy (1979:19) defined the marking scheme as:

A marking scheme is a comprehensive document indicating the explicit criteria against which candidates' answers will be judged: it enables the examiner to relate particular marks to answers of specified quality.

In assessing the oral proficiency, it is necessary to provide comprehensive descriptions of levels of performance to aid the examiners in making necessarily subjective judgements about the worth of candidates' answers. The marking scheme as described by Murphy (ibid: 14) is used to:

- to assist the examiner and those who will mark the paper to check the content validity of the tasks which are being set;
- to help the testers to check that the demands made in the examination are appropriate and in accordance with stated aims and with stated aims and objectives;

- to ensure that, where there is more than one examiner, each examiner marks in exactly the same way, awarding equal marks for equal levels of performance;
- to ensure that each examiner marks consistently throughout the marking period.

Murphy adds saying that in assessing oral proficiency it is essential that tasks and marking schemes are subjected to a rigorous process of moderation before they become operational.

At the same time as the activities are clearly designed the assessor should consider the appropriateness of the marking scheme. The following are examples of the types of questions assessors might address themselves to (Murphy 1979: 61):

- Does the marking scheme anticipate responses of a kind that candidates are likely to make?
- Are the marks allocated to each task commensurate with the demands that a task makes on the candidate?
- Does the marking scheme indicate clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable?
 - Does the marking scheme allow for possible alternative answers?
 - Has the marking scheme reduced to the minimum possible, the amount of computational work which the examiner has to undertake to finalise a mark for a candidate's performance?
- Does the marking scheme, by specifying performance criteria, reduce as far as possible the element of subjective judgement that the examiner has to exercise in evaluating candidates' answers?

- Are the abilities being rewarded those which the activities are designed to assess?
- Can the marking scheme be easily interpreted by a number of different examiners in a way which will ensure that all mark to the same standard? The Marks allocated can reflect the group and individual effort.

-Group marks will be given for:

Organization:

- Teamwork.
- Structure of presentations.
- Organization of ideas.
- Timing.

Content:

- Research.
- Ideas.
- Argument.
- Approachability.

Interest:

- Presentation.
- Variety.
- Awareness of audience.

- Individual marks will be given primarily for:

Language:

- Pronunciation
- Grammatical accuracy.

- Range vocabulary, grammatical structures etc.

Communicative quality

- Fluency, speed etc
- Ability to

communicate ideas But also:

Content:

- Research.
- Ideas.
- Argument.
- Approachability.

Organization and timing

Interest:

- Presentation.
- Variety.

Francis (1981:21) reports little difference in reliability between impression marks and an analytic grid system. However, studies showed that many examiners involved in different boards of testing preferred the grid system, which they felt to have greater validity, as it compels marks to be given "for discrete and specific skills within the sub-tests". An analytic marking scheme also forces examiners to give careful consideration to the list of skills to be tested and to their relative weighting. This list of skills can, of course, vary according to the level of the student and the type of course. Without a clear definition of what we are looking for in an oral examination, it is very difficult, if not impossible, to achieve reliable results.

5- The Assessment Scales

In order to assess the communicative ability, two schemes have been suggested: ACTFL- the American Council on the Teaching of Foreign Languages - is expressed in general terms to cover any foreign language, and Carroll's Band of general assessment. Both have nine categories which are more or less parallel. The correspondence indicated here is approximate. It could be argued that Carroll's band 9 - "Expert," for example, is apparently more advanced than the ACTFL "Superior" category. We believe that one of the major advantages of this kind of assessment scale is that it is relatively easy to utilize, as it provides a clear description of communicative ability. Even where not implemented rigorously, it provides some reference points. It is true that with some training, this scale can be quite reliable with similar results from one tester to another.

Table 2: Superior - Expert user

ACTFL - American Council on the Teaching of Foreign Languages		Carroll, B (1980)
<p>The Superior level is characterized by the speaker's ability to:</p> <p>1) participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics; and</p> <p>2) support opinions and hypothesize using native-like discourse strategies.</p>	<p>Superior Able to speak the language with sufficient accuracy to participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics. Can discuss special fields of competence and interest with ease. Can support opinion and hypothesize, but may not be able to tailor language to audience or discuss in depth highly abstract or unfamiliar topics. Usually the Superior level speaker is only partially familiar with regional or other dialectical variants.</p> <p>The Superior level speaker commands a wide variety of interactive strategies and shows good awareness of discourse strategies. The latter involves the ability to distinguish main ideas from supporting information through syntactic, lexical and suprasegmental features (pitch, stress, intonation). Sporadic errors may occur, particularly in low frequency structures and some complex high-frequency structures more common to formal writing, but no patterns of error are evident. Errors do not disturb the native speaker or interfere with communication.</p>	<p>9 - Expert user</p> <p>Communicates with authority, accuracy, and style. Completely at home in idiomatic and specialist English</p>

Table 3: Advanced Level

ACTFL - American Council on the Teaching of Foreign Languages		Carroll, B (1980)
<p>The advanced level is characterized by the speaker's ability to:</p> <p>1) converse in a clearly participatory fashion;</p> <p>2) initiate, sustain, and bring to closure a wide variety of communicative tasks, including those that require an increased ability to convey meaning with diverse language strategies due to a complication or an unforeseen turn of events;</p>	<p>Advanced-Plus: Able to satisfy the requirements of a broad variety of every day, school, and work situations. Can discuss concrete topics relating to particular interests and special fields of competence.</p> <p>There is emerging evidence of ability to support opinions, explain in detail, and hypothesize.</p> <p>The Advanced-Plus speaker often shows a well developed ability to compensate for an imperfect grasp of some forms with confident use of communicative strategies, such as paraphrasing and circumlocution.</p> <p>Differentiated vocabulary and intonation are effectively used to communicate fine shades of meaning.</p> <p>The Advanced-Plus speaker often shows remarkable fluency and ease of speech but under the demands of Superior-level, complex tasks, language may break down or prove inadequate.</p>	<p>8 - Very good user</p> <p>Presentation of subject clear and logical with fair style and appreciation of attitudinal markers. Often approaching bilingual competence.</p>

<p>3) Satisfy the requirements of school and work situations; and</p> <p>4) narrate and describe with paragraph-length connected discourse.</p>	<p>Advanced Able to satisfy the requirements of everyday situations and routine school and work requirements. Can handle with confidence but not with facility complicated tasks and social situations, such as elaborating, complaining, and apologizing. Can narrate and describe with some details, linking sentences together smoothly. Can communicate facts and talk casually about topics of current public and personal interest, using general vocabulary.</p> <p>Shortcomings can often be smoothed over by communicative strategies, such as pause fillers, stalling devices, and different rates of speech. Circumlocution which arises from vocabulary or syntactic limitations very often is quite successful, though some groping for words may still be evident. The Advanced-level speaker can be understood without difficulty by native interlocutors.</p>	<p>7 - Good user:</p> <p>Would cope in most situations in an English-speaking environment. Occasional slips and restrictions of language will not impede communication</p>
---	--	---

Table 4: Intermediate Level

<p>ACTFL - American Council on the Teaching of Foreign Languages</p>		<p>Carroll, B (1980)</p>
<p>The Intermediate level is characterized by the speaker's ability to:</p> <ol style="list-style-type: none"> 1) create with the language by combining and recombining learned elements, though primarily in a reactive mode; 2) initiate, minimally sustain, and close in a simple way basic communicative tasks; and 3) ask and answer questions. 	<p>Intermediate-High: Able to handle successfully most uncomplicated communicative tasks and social situations. Can initiate, sustain, and close a general conversation with a number of strategies appropriate to a range of circumstances and topics, but errors are evident. Limited vocabulary still necessitates hesitation and may bring about slightly unexpected circumlocution.</p> <p>There is emerging evidence of connected discourse, particularly for simple narration and/or description.</p> <p>The Intermediate-High speaker can generally be understood even by interlocutors not accustomed to dealing with speakers at this level, but repetition may still be required.</p>	<p>6 - Competent user</p> <p>Although coping well in most situations he is likely to meet, is somewhat deficient in fluency and accuracy and will have occasional misunderstandings or significant errors.</p>

	<p>Intermediate-Mid: Able to handle successfully a variety of uncomplicated, basic and communicative tasks and social situations. Can talk simply about self and family members.</p> <p>Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure time but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms.</p> <p>Pronunciation may continue to be strongly influenced by first language and fluency may still be strained.</p> <p>Although misunderstandings still arise, the Intermediate-Mid speaker can generally be understood by sympathetic interlocutors.</p>	<p>5 - Modest user:</p> <p>Although manages in general to communicate, often uses inaccurate or inappropriate language</p>
	<p>Intermediate-Low Able to handle successfully a limited number of interactive, task-oriented and social</p>	<p>4 - Marginal user:</p> <p>Lacking in style, fluency and accuracy,</p>

	<p>situations. Can ask and answer questions, initiate and respond to simple statements and maintain face-to-face conversation, although in a highly restricted manner and with much linguistic inaccuracy. Within these limitations, can perform such tasks as introducing self, ordering a meal, asking directions, and making purchases. Vocabulary is adequate to express only the most elementary needs. Strong interference from native language may occur.</p> <p>Misunderstandings frequently arise, but with repetition, the Intermediate-Low speaker can generally be understood by sympathetic interlocutors.</p>	<p>is not easy to communicate with; accent and usage cause misunderstandings.</p> <p>Generally can get by without serious breakdowns.</p>
--	---	---

Table 5: Novice

ACTFL - American Council on the Teaching of Foreign Languages		Carroll, B (1980)
<p>The Novice(Novice level is characterized by the ability to communicate minimally with learned material.</p>	<p>-High: Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombination of their elements. Can ask questions or make statements involving learned material. Shows signs of spontaneity although this falls short of real autonomy of expression. Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. Pronunciation may still be strongly influenced by first language. Errors are frequent and, in spite of repetition, some Novice-High speakers will have difficulty being understood even by sympathetic interlocutors.</p>	<p>3 - Extremely limited user: Does not have a working knowledge of the language for day to day purposes, but better than an absolute beginner. Neither productive nor receptive skills allow continuous communication.</p>
	<p>Novice-Mid: Oral production continues to consist of isolated words and learned phrases within very predictable areas of need, although quality is increased.</p>	<p>2 - Intermittent user: Performance below level of day to day knowledge of the language.</p>

	<p>Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words. Speaker may have some difficulty producing even the simplest utterances. Some Novice-Mid speakers will be understood only with great difficulty.</p>	<p>Communication occurs only sporadically.</p>
	<p>Novice-Low: Oral production consists of isolated words and perhaps a few high-frequency phrases. Essentially no functional communicative ability.</p>	<p>1/0 - Non-user: May not even recognize with certainty what language is being used.</p>

III-Procedures in Testing Oral Proficiency in the Department of English

In this section, we will describe some of the issues associated with oral proficiency testing in large classrooms, in the Department of English, Faculty of Languages and Human Sciences, University of Béjaia.

At the end of each semester, teachers are faced with the necessity of assessing what their students have learned and representing this by a grade. To convert many hours of instruction into a single letter or number is a daunting task. Ideally, tests should endeavour to help not only teachers or administrators, but also aid students in assessing their performance. Tests should serve both evaluative and educational functions. Through our observation and investigation, we realized that many of the tests are designed primarily to dispense grades. Typically, students take a test at the end of each semester and receive a grade with no comments a few weeks later. Such feedback is of marginal value. Prompt feedback is crucial since most students tend to forget the details of their tests soon after completing them. Unless feedback is specific and immediate, its pedagogical value is very limited.

Rather than waiting until a battery of questions has been completed before providing feedback, we think that we need to provide suggested answers soon after each question has been raised. With the oral examination procedure described in this study, prompt and point-by-point feedback can be offered without concern of students manipulating test results. A student whose grade in a class is based on merely a few test results can justifiably feel his / her skill has not been adequately measured. In such a case, the chance of an inflated or deflated score is undoubtedly high.

A leading concern many teachers have about any test is the relative ease of administration. With up to forty students or more in a class, few teachers look forward to devoting the hours necessary to assess the performance of their students individually. This leads to an interesting irony: although the best way to assess the oral proficiency may be verbally, the time involved in administering individual oral exams has prompted many teachers to opt for more convenient testing formats. Perhaps the easiest test to administer is a written examination in a multiple choice or cloze format. However, the extent that such examinations can accurately assess oral proficiency is most of the time questionable. We have realized that some shy students who never speak in an actual conversation often manage to fill in the right blanks on a test sheet.

Besides final examinations, some teachers base their final grades on quantifiable aspects of classroom performance such as attendance, completed homework, or the cumulative results of mini-quizzes. We believe and regard this as a valid way of measuring classroom performance; it is not a valid measure of English ability. A student who attends class faithfully, turns in the recommended assignments and makes fledgling attempts to speak might receive a higher grade than one who is occasionally absent - even though the latter student may in fact have better oral skills.

Another aspect which we have noticed while meeting students is the anxiety many of them have about testing. It would not be an understatement that most met students have "test phobia." Teachers are also likely to have mixed feelings about the testing process. We also noted that, there was no program-wide oral assessment. Teachers created their own in-class processes to assess speaking and listening or, more often, based their evaluations entirely on classroom observation. The major limitation to such evaluation is related to the lack of trained and qualified teachers.

The traditional tools used for oral proficiency, measure language usage, not language use, that is, they tell what the student knows about the operation of the language but not how the student performs when producing the language. That is why; it is useful to see how and how much our students can express their ideas with someone other than the teacher. Sometimes, they can do better. It reminds us that our students need to communicate with other people in a different context. It is more realistic than the classroom."

Conclusion

We can conclude by saying that assessing the oral proficiency of our learners is indeed a complex task. In order to avoid the common difficulties encountered while testing such a skill, there is a need to determine the criteria to be considered; such as validity and reliability. We recognise that some of the procedures of testing oral proficiency in the Department of English and their implication on the process of learning in general need to be reconsidered. For this purpose, and in order to improve the reliability of oral examinations, two schemes have been suggested: ACTFL- the American Council on the Teaching of Foreign Languages - is expressed in general terms to cover any foreign language, and Carroll's Band of general assessment.

CHAPTER FIVE

ORAL PROFICIENCY TESTS

Introduction	131
I- Traditional Assessment Techniques	131
1- Multiple Choice Questions	131
2- Oral Interview	135
3- Group Work	138
4- Problem-Solving	140
5- Role Play	142
6- Discussion	144
II-Self Assessment	146
1-Definition	146
2-Research in Self-Assessment	147
3-Aspects of Self-Assessment	150
3.1-Responsibility	151
3.2-Learner Training	152
3.3 Innovation in Self-Assessment?	153
4-Reservations as to the Use of Self-Assessment: The Problem of Bias	154
III-Portfolio Assessment	156
1-What is a Portfolio	156
2-Why Use Portfolio Assessment	157
3-Stages in Implementing Portfolio Assessment	159
4-Advantages and Disadvantages of Using Portfolio Assessment	164
Conclusion	167

Introduction

Testing oral proficiency remains widely regarded as the most challenging of all language exams to prepare, administer and score. Investigations and research have shown that in very recent years, there has been a growing interest in the application of testing procedures that are totally different from traditional forms of assessment. More authentic forms of measurement have become increasingly popular in the foreign language classes. These forms of assessment are more student centered in that, in addition to being an assessment tool, they provide the student with a tool to be more involved in their learning, and give them a better sense of control of their own learning. In this chapter, we will illustrate some of the commonly used oral proficiency tests in different educational institutions.

I-Traditional Assessment Techniques

1-Multiple Choice Questions

The multiple-choice question (MCQ) test is nowadays widely regarded as being one of the most useful of all objective item types. Although it is amongst the most difficult of all objective tests to construct, it is simple to score and administer.

We believe that the following general principles should be observed when (MCQ) are constructed:

- Each multiple-choice item should have only *one* answer. This answer must be absolutely correct, unless the instruction specifies choosing the *best* option (as in some vocabulary tests).
- Only one feature at a time should be tested because it is usually less confusing for the testee and it helps to reinforce a particular teaching point.

- All MC items should be at an appropriate level to the linguistic ability of the testees.

- MC items should be as brief and as clear as possible.

The optimum number of alternatives, or options, for each MC item is said to be five in most tests. Although a larger number, say seven, would reduce even further the element of giving the right answer by chance, it is extremely difficult and often impossible to construct as many as seven good options. Indeed, since it is often very difficult to construct items with even five options, four options are recommended for most classroom tests. Many writers recommend using four options for grammar items, but five for vocabulary.

Before constructing any test item, the test writer must first determine the actual areas to be covered by MC items and the number of items to be included in the test. Too long tests are undesirable because of the mental strain and tension which may be caused among the students taking the test. The number of items included in a test will vary according to the level of difficulty, the nature of the areas being tested, and the purpose of the test. The teacher's own experience will generally determine the length of a test for classroom use.

The initial part of each MC item is known as the *stem*; it is followed by options/responses/alternatives; one option is the answer, correct option or *key*, while the other options are distractors. The task of a distractor is to distract the majority of poor students (i.e. those who do not know the answer) from the correct option.

- The stem:

The primary purpose of the stem is to present the problem clearly and concisely. The testee should be able to obtain from the stem a very general

idea of the problem and the answer required. At the same time, the stem should not contain extraneous information or irrelevant clues, thereby confusing the problem being tested. It should convey enough information to indicate the basis on which the correct option should be selected.

- The correct option

For normal purposes of testing, this should be clearly the *correct* or *best* option. Thus, it is most important that each item is checked by another person to make sure not to confuse the student by having a different number of correct options for each item.

- The distractors:

Each distractor, or incorrect option, should be reasonably attractive and plausible. It should *appear* right to any testee who is unsure of the correct option. Items should be constructed in such a way that students obtain the correct option by direct selection rather than by the elimination of obviously incorrect options. Each distractor should be grammatically correct when it stands by itself; otherwise, testees will be exposed to incorrect forms. Unless a distractor is attractive to the student who is not sure of the correct answer, its inclusion in a test item is superfluous. Plausible distractors are best based on:

- mistakes in the student's own written work,
- their answers in previous tests,
- the teacher's experience,
- and a contrastive analysis between the native and target the language.

Distractors should not be too difficult nor demand a higher proficiency in the language than the correct option. If they are too difficult, they will only succeed in distracting the good student who will be led into considering the correct

option too easy (and a trap); there is a tendency for this to happen, particularly in vocabulary test items.

Moreover, unless there are good reasons for weighting different items (using the appropriate statistical methods), it is advisable to award equal marks for each item.

MCQ tests are frequently criticised on the grounds that they are simpler to answer than subjective examinations. Items in an objective test, however, can be made just as easy or as difficult as the test constructor wishes. In devising such a test, the constructor does not only select and construct his/her items carefully but analyses students' performance on each item and rewrites where necessary so that the final version of his test is highly likely to discriminate between the students.

Another criticism is that objective tests of the MC type encourage guessing. However, 4 or 5 alternatives for each item are sufficient to reduce the possibility of guessing. Furthermore, experience shows that candidates rarely make wild guesses: most base their guesses on partial knowledge.

The fact that a test is apparently completely objective — the MCQ test for example, gives no guarantee that it will be a good test. It will be a very poor test if:

- the wrong features of the target language are tested;
- -irrelevant areas are emphasised in the test simply because they are "testable"; and-the test items are poorly written.

It should never be claimed that objective tests can do those tasks which they are not intended to do. They can never test ability to *communicate* in the target language, nor can they evaluate actual performance. A good classroom test will probably contain both subjective and objective test items.

2-The Oral Interview

Interviews are an oral data collection technique that captures information verbally. They are widely used for assessing second language learners' speaking proficiency to sustain a conversation. They are generally structured so that interviewees are given a series of tasks/situations to which they have to respond. According to Hymes,(1989:2-3), the speaker reacts to conversational initiatives of the conversational partner, primarily with formulaic or rote utterances in words and phrases, such as greetings and often- heard expressions, and lists of vocabulary items that many have been learned in groups, such as colors, numbers, days of the week, months, articles of clothing.

In the classroom, teachers can look at three types of interview formats: individual, peer, and group.

- **An Individual interview** is a face-to-face spoken interaction between two participants: one is an expert who conducts the interview, and the other the learner. This type of interview, while time consuming, provides the most in depth information about the student. Students have individual opportunities to talk about their learning or to perform in the language and, respond positively to the one-on-one attention. There are specific guidelines to follow for an individual interview to be successful. First, the interview can only be useful if we begin by determining a purpose. Once the purpose is determined, we prepare a set of activities which will allow the students to demonstrate their oral proficiency, for example, how to show a way for someone on a map. Next, we prepare questions for the interview. It is important for validity and reliability of the results that a set of questions go along with the activities and that crucial questions are asked to all the students. The interview should be conducted periodically, thereafter to monitor students' progress, and to allow the teacher to

make an individualized diagnosis and/or a class-wide diagnosis and refocus teaching to meet the students' needs.

- **Peer interviews** consist of two students, one as the interviewee and the other as the interviewer. Students can take turns at each of the roles. These interviews are easier for the teacher to manage because students are collecting the information.

- **Group interviews** are generally retrospective in that students are asked what they think they would do in a given situation. For instance, a group may be asked questions such as "How do you remember vocabulary? What advice would you give to a beginning student on how to learn vocabulary? Group interviews can give a picture of trends in a class or reveal areas of interest in which individual interviews could be used to follow up on.

To strengthen the reliability of the data collected through the interview, an interview script needs to be developed, piloted, and followed by all interviewers for each interview conducted. In addition to the interview script, a method of recording the oral information needs to be chosen, for example audio tape recorders and video cameras. The information collected through interviews generally refers to students' proficiency levels, and learning motivations, metacognitive awareness, learning strategies use, self-efficacy levels, learning preferences and styles. It has practical classroom applications beyond formal research purposes: it can be used to inform instruction, to make assessment decisions, and to help learners become more aware of their learning.

Conducting successful interviews requires that many issues are considered and discussed at length with students in the context of developing interview skills. The interview should:

- Let the student talk and to listen very attentively to what is being said. In order to encourage the interviewee to talk, it is useful to ask open-ended questions such as: What are you thinking? How do you know that?
- Start with easy-to-answer, non-controversial question: the interviewee may need time to think. Interviewers must remember never to interrupt a good story to ask for details. The interviewer can always ask follow-up questions afterwards to clarify something.
- Take notes the tape recorder might malfunction and watch for opportunities to expand on topics.
- The interviewer should watch body language for signs of fatigue or discomfort:
- Are the questions being asked too personal or too painful for the interviewee to answer?
- Is the interviewee too exhausted to continue? Sometimes it is wise to stop for a break or make arrangements to continue another day.

If the interviewer uses gestures, the student should verify them verbally. The interviewer should take care not to sound judgmental, impatient or disrespectful at any time during the interview.

Finally, we can say that it is most helpful if the teacher can give the students clear, specific guidelines about what needs to be done. In this situation, the teacher becomes the facilitator of the learning process instead of the giver of information.

3. Group Work

A group work test is a classroom situation where the students function within smaller units (groups). In these small units, an atmosphere conducive to oral exchanges is created, and the possibilities for interpersonal interactions are multiplied. For example, students listen to or read a text of five paragraphs. They are then divided into two groups. Each group selects a paragraph of the text just read or listened to and prepares to answer the questions asked by the other group. Each group has to work collectively to scan a paragraph for detailed comprehension and formulate questions to test the comprehension of the other group. The preliminary work involves oral interaction asking and answering questions, paraphrasing some ideas of the paragraph, suggesting questions, agreeing, disagreeing, and proposing modifications. After some minutes preparation, the exchange between groups starts. The "confrontation" which follows is genuine because of the information gap created between the two groups: neither group knows the questions which will be asked nor the answer.

Oral interaction based on a genuine attempt to find a collective solution to a problem is at the basis of group work. Group work is a meaningful testing technique because the students need to interact with their peers and to develop a range of social and linguistic skills which are of great importance in terms of social relations. For this to happen, students should have some familiarity with the discussion topic prior to the discussion. Ideally, we think they should be exposed to a variety of discussion stimulants, including new vocabulary in context which would have been presented in the form of drills. The primary teacher's role is to stimulate thinking about the topic of discussion in advance, i. e. before the group work test starts. Therefore, the teacher remains essentially out of the discussion, or at least participates only to the extent that his/her presence does not destroy the conversational quality of the group. If the teacher

feels he/she must continually correct students' mistakes, he/she will certainly succeed in destroying the cohesiveness, therefore the conversational quality of the group.

Group work technique, communicatively contrived, presents the following advantages.

- The members of a group engage in "authentic negotiation"(Breen, 1997) in an attempt to solve a problem. They do not simply throw words at each other; they interact orally with a purpose.
- Spontaneous speech at a comfortable speed is enhanced in a student and
- During the oral interaction, there is a continuous process of interpreting and evaluating utterances as well as expressing views. It also has some disadvantages, like:

- it may kill the spirit of self reliance,
- and • from the student's point of view, the value of help from the colleagues is less than the teacher's. Group work involving communicative tasks is essential for promoting oral communication because it creates a situation where learners learn with friends and from friends .It favors group seating arrangement. It drastically modifies the role of the teacher and his/her conduct of the class; however, it should be mentioned that to eliminate the domineering position of a teacher is not to invite chaos.

As a conclusion, we can say that group work is important in that it demands maximum students' participation in an orally purposeful activity.

4-Problem-Solving

In problem solving, learners are involved in interpreting a task and in developing it through oral expression and negotiation. Problem solving can either be on an individual or collective. The latter, calling for cooperative negotiation, is most rewarding in oral language terms; it can be of two kinds: short term tasks and projects. The former can be completed in the course of one class session (short term) while the latter spans many sessions and more preparatory research.

An example of a short term problem-solving activity consists in presenting the students with a text which has been deliberately left unfinished and requiring them to provide an end, a conclusion. Another example is presenting the students with recordings they have to listen to, and asking them to identify the speakers, the place, and the situation, to give a short oral commentary. Students may also be presented with the "for" side of an argument and asked collectively to constitute the "against" side of it. These short term activities are task-centered and can be achieved within one class session. For a project, in groups, students are asked to design and make an oral presentation of, for example various new models of business documents (vouchers, cheques, receipts books, ledgers...) stating the modification from past documents and outlining their advantages, or to search into the problems of national companies, and present their findings to a class. These examples show that a project is more time consuming than a short term task because it requires more information search, information exchange and information synthesis.

In both short term and project tasks, and the situational language use is centered on a problem, situations created are conducive to interacting orally and collaborating in order to accomplish the task.. To create such a situation, the

selection of activities should not be haphazardly done. It should take into consideration the level of the learners, the scope they provide for oral language use, their relevance to the students, their profession and the language skills to be practised. Problem-solving is an appropriate device for testing oral proficiency because:

- students are fully involved, and there is an immediate pay off in completing the task, the pay off being a motivating factor;
 - self-expression is enhanced through exposés (projects), joint efforts to find solutions to problems and the general discussion during and after each presentation;
 - students are put in a dependency situation where they have to communicate orally with group members during the search for exchange of relevant information on the subject;
- and • during the oral presentation of findings, student colleagues interrupt with questions, comments, and points of observation, thereby creating a truly oral communicative scene, void of the recitational aspect of pattern drills.

Nevertheless, the problem solving technique has its limitations:

- in group tasks, it is difficult to efficiently supervise various activities or projects at the same time.

As a final word about problem solving, we would say that the assessment technique is innovative in that it is the search for the unknown. The unpredictability of the outcome fosters curiosity in learners, and the oral interactive atmosphere created helps in developing oral proficiency.

5- Role Play

Role play is a good technique to measure interaction in foreign language classes. It has become a standard technique to test interaction in foreign language classes, and in teacher and management training. It represents a shift of emphasis to more realistic conversation and communication. The activity ranges from telling a story to acting out a situation where participants have to improvise and create. Role play encourages interaction among the learners as they relate to each other through their new identified roles.

Revell (1979:60) sees role play as: "an individual's spontaneous behavior reacting to others in a hypothetical situation." This implies that a role play participant puts a fictitious identity in an imagined situation to present the views of a person without necessarily sharing them; for example, "x" has applied for a position of a salesman and has been called up for an interview before a selection board..

Mugglestone (1977) states that each learner in role playing comes out under the four following categories:

- that of acting out a role already performed in the L 1 and that needs to be performed in English, for example being a guest or host at a party;
 - that of acting out a role that already performed in the L but which is unlikely to be performed in English, for example being a husband or a wife;
 - that of acting out a role he has not performed in his L 1 but that he will need to perform in English, for example being a student, or about to become a post-graduate at a British university, and needing to participate in tutorials and seminars;
- and • that of acting out a role not performed in the L 1 and which is unlikely to be performed in English, for example, being a policeman.

Robinson (1981) suggests that role play may take many forms, but in essence, it is an improvisation where the participants build up their own characters, talk, movements, situation, structures, themes and messages. Those who are not taking role (the rest of the class and the teacher) can intervene to find out what is going on and give help, either with the role, or with the language they can replace, or join the role players from time to time.

From the oral communication point of view, role play presents several advantages.

- As a direct interactive method, role play encourages and promotes spontaneous oral exchanges between participants. There is an active participation on the part of the students: they contribute to interact with the other participants instead of reciting already memorized stretches. Indeed, as Dickson (1981:382) puts it: "Learners say what they want to say and not what someone has told them to say".

- During the communicative transaction, both the learner's language fluency and knowledge of syntax are tested, although the emphasis is more on the former.

- The interaction involves, almost exclusively, the learners alone-the teacher is there to maintain a non intrusive role-this means an increase in individual student's speaking time.

- Role players learn the interactive skills of arguing, informing, persuading, complaining, discussing, compromising... .

and • Role play is a challenging activity: being a dynamic technique, it deals with constantly changing situations as each participant "shifts into" the discussion to add a new element.

This assessment technique also has some disadvantages.

- A persistently inarticulate and shy student who does not play his/her role/part appropriately destroys the whole framework, as there is not much interreliance.

and • The teacher is not there to serve as a guideline or example, but he/she can occasionally be one of the players.

Opting for role play means reexamining and possibly revising one's style of testing and teaching. The pedagogical implications of role play are not different from those of the other techniques already mentioned. It however suffices to recall that these considerations center around class organization.

6- Discussion

A discussion as a technique for measuring oral expression aims to give and receive opinion, to give and share points of view. For example, after the presentation of a short exposé where the class can resort to commenting on the subject, with learners expressing their views. It is thus a learning experience which in itself helps in promoting the use of oral expression in a genuinely interactive situation.

The rationale behind a discussion is that when learners are presented with a relevant problematic situation as open-ended as possible, they interact orally in a collective effort to resolve the task. In a discussion, not only the ability to speak is fostered, but also the ability to listen and understand, to answer appropriately and intervene efficiently. A discussion class is intended to provide an open forum for learners to air their views orally about certain issues, to learn from others, to inform and be informed on certain matters of topical interest, professional, educational, social and otherwise.

We have to keep in mind that topics for discussion are not selected at random. Some of the guidelines for selection include:

- relevance to the professional/educational and age group needs;
 - specifying the educational level of the students and the level of expert knowledge required for students to feel comfortable with the topic;
- and - a variety of sources of input materials, newspapers, radio, video-recording, or simply a text short enough to enable learners to concentrate on discussion.

The advantages of discussion are not different from those of the other techniques already discussed, but the following are worthy of a special note.

- Participants are engaged in interpreting utterances and responding appropriately on the spur of the moment.
 - The learners initiate their own language and put it to communicative use without having to repeat pre-rehearsed stretches.
 - There is a give and take of information to an extent that a great deal of information is available on the subject at the end of the session.
- and • There is a scope for further communicative use of a discussion session as when a group leader is asked during a "report back" session to recapitulate the main trends of arguments that emerged, or give a summary of the proceedings of the last discussion class, or when a taped debate/ discussion is played back in a different class for listening commentary.

Discussion, nonetheless presents some disadvantages.

- The technicality of some topics preferred by the students may be a handicap to the teacher, although some specialists can be invited to participate in such discussion.

- There may be a tendency to "talk away" the time, with nothing effective added to the student's knowledge. In some discussion, there is a great difficulty in avoiding "yes/ no" and "I don't know" responses likely to close down the discussion. This disadvantage can be overcome by not considering discussion as time filler but as a session to be planned, with the students prepared in advance for debates, points for and against could first be done individually as a homework, and one of the students can act as a chairperson.

As a final word, it could be mentioned that discussion should lead to oral reporting, essay writing, summary writing; it develops not only speaking fluency, but also its "associates": listening, comprehending, interpreting utterances, responding appropriately.

II-Self-Assessment

1-Definition

Self-assessment relates to learners' evaluation of their own-or their peers-home or classroom work, usually based on guidelines provided by the teacher. These take the form of checklists, questionnaires, diary or journal reports and all sorts of exercises available to learners at self-access centres. Learner-evaluation, in this sense, is concerned with feedback and self-improvement and is distinct from evaluation in a formal sense, i.e., for the purposes of assigning grades. The possible benefits of self-Assessment are discussed in relation to its potential to:

- increase learners' responsibility for their own learning,
- alleviate the burden of assessment imposed on teachers,
- develop learners' study skills, thus extending the boundaries and objectives of the English class beyond the limits of a foreign language teaching course.

Both the rationale and techniques which can be employed in self-assessment are outlined, drawing upon the available literature on the subject and projecting into the context of the first year students at the Department of English.

2-Research in Self-Assessment

Research findings in the area of self-assessment (Oskarsson 1984 for a review) seem to point to a fairly consistent correlation between self-assessment and other forms of external (i.e., from outside the classroom or school) measures or teachers' assessment. Most of the studies reported in the literature have involved adult students in the context of university courses. These studies are marked by the specificity of the situation in which they are embedded by the aims or views of each researcher. A look at some of these studies might, however, give us an indication as to the possible applications of self-assessment in the context of foreign language teaching to first year learners in our university.

- **Oskarsson:** The most extensive work in the area of self assessment is probably that of Oskarsson, from the University of Göteborg, in Sweden (Oskarsson 1978, 1984, 1989). He conducted a pioneering study for the Council of Europe, published in 1978, in which possible forms of self- assessment were discussed and related research projects were described.

Oskarsson's first review of the literature (in the late 70 s) revealed that self-assessment was being conducted mainly through the use of discrete-point tests of vocabulary, grammar and reading comprehension. Very little was done in relation to assessment of oral skills or the use of other instrument of oral skills

or in relation to the use of other instruments for self-assessment — like questionnaires or peer assessment of simulated real- life situations.

Field experiments of the ideas proposed then generally showed high correlation between self-assessment and external tests. These studies also confirmed an expected need for learner training in relation to assessment.

- **Bachman and Palmer (1981)** - In this study, the authors used self-assessment as part of a multitrait- multimethod investigation into the construct validity of tests of speaking and reading. The purpose of the investigation was to find out whether speaking and reading abilities are independently measurable, i.e., whether they are in fact separate constructs.

Their self-assessment instrument consisted of written questionnaires in the subjects' mother tongue (Mandarin Chinese). The questions tried to elicit subjects' perceptions of their general language ability as well as their control of linguistic forms, for example:

Q. "How hard is it for you to use different kinds of English with different kinds of people?"

A. "Impossible/ Very hard/ Not very hard/ Very easy."

Q. "How many different kinds of pronunciation mistakes do you make in English? "

A. "I make pronunciation mistakes in almost everything/ many kinds/ only a few kinds/ I almost never make pronunciation mistakes."

The results obtained from the questionnaires were correlated with the results of an oral interview and an oral translation method (considered by the researchers to be the best measures of speaking and reading skills, respectively). Bachman and Palmer found that the reliability measures were as high as 85-.96 for the self-assessment method. Substantial intercorrelations

between self-rating and the other two methods were obtained, with coefficients ranging from .51 to .74.

- **Le Blanc and Painchaud (1985):** This study used self-assessment for the purpose of placement. It was conducted at the University of Ottawa where new students have to be assessed in their proficiency of either English or French as a second language.

Questionnaires covering general ability in the four skills-with questions similar to those used by Bachman and Palmer were administered and correlated with proficiency tests of the same skills. Results showed not only that there were high correlations between the different measures, but also that after some time the percentage of changes from initial placement was very low. Self-assessment questionnaires were also found to be much easier and quicker to administer than conventional tests.

- **Von Eiek (1985):** The test developed by Von Eiek, from the University of Gothenburg, Sweden, is one of the most important ones in the area. It is an instrument devised for self-diagnostic purposes to be used by adult migrants studying Swedish as a second language.

The test battery consists of 1,500 items covering six areas: the four skills plus grammar and vocabulary. The questions are all formulated in the target language, and the student has a three choice answer form with the options: "yes, absolutely"; "I think so"; or "no". Examples of questions are:

- "Vocabulary": Do you know the meaning of the underlined word below? -
- Listening: Do you know which picture the speaker is referring to? "

For each of the six areas, there are 25 questions, each one providing 10 sub- questions of different degrees of difficulty. The results of such a questionnaire provide learners with a self-rated profile that can be used for

diagnostic purposes and some sort of measurement of level of proficiency in the various areas covered.

- **Eurocentre Bournemouth (in Holec 1988; Oskarsson 1984)** the work done at the Eurocentre language school in Bournemouth has been directed at continuous self-assessment of classroom activities as a means to improving the learning process.

Different ways of self-assessment have been used, such as laboratory follow-up to oral practice lessons, with recording of students' interactions being commented on by the teacher, after which a new recording is made and compared by the students with the previous version of progress videos in which students are video taped doing short role-plays at regular intervals (Ferris, in Holec 1980). Towards the end of the term the students are shown the films and asked to evaluate their performance on the basis of a form covering the areas of fluency, pronunciation, grammar and general impression.

These approaches are reported by participants as being highly valuable in promoting learner motivation and autonomy. There is also, according to them, a usually good index of agreement between learners and teachers ratings.

3-Aspects of Self-Assessment

Implementing a self-assessment project will certainly require the allocation of a good deal of time in any course programme. Familiarising students with the idea of self-assessment and training them to do so is not something one can do with a five minutes chat at the end of a class session.

Even considering that one of the aims of self- directed learning is to enable learners to work on their own outside the classroom hours, one still has to

devote some of the class time to promote this independence. We believe that planning a self-assessment project should include considerations about the syllabus to be covered within a certain period of time. Self —directed learning which is included in the new reforms of higher education (within LMD system) should lead to more effective and faster learning, in the long run. However, we realize that such a project is extremely time consuming and would probably require the reduction of the input load in terms of the contents of a given syllabus.

3.1-Responsibility

The term "responsibility", when applied to the teaching/ learning situation, can be a tricky one. On the one hand, it is desirable to engage learners actively in the educational process, so that whatever is taught can be learnt, or so that "input" can more easily become "intake". In this sense, we suppose we could use the term responsibility to refer to the learners' share in the process of education. This does not imply that teachers or institutions lose their own share in conducting the process although a learner- centred orientation in education, as Allwright points out (1979) does imply abdication of responsibility.

This view, however, may not be very clear to students, especially if their previous educational experiences have been marked by conservative practices in which the roles of teachers and learners are very distinctly defined. These learners might resent being given the task (the chance?) of self- assessment. Imposing it on such a group would then be far from the objectives which self- evaluation seeks to promote.

This possibility of some resistance on the part of students in relation to self- assessment has to do with the power with which institutions and teachers

are invested in relation to deciding what is supposedly best for students. Self-assessment will only represent an innovation in educational terms if it is a way to empower learners to make decisions about themselves. Otherwise, it will be just another gimmick to fill the pages of teachers' manuals.

3.2-Learner Training

The aspect of learner training seems to us to relate to three areas.

- Firstly, it has to do with acquainting learners with the concept of self-assessment, i.e., what it entails, what it seeks to promote and how it is conducted.

- Secondly, it relates to study skills and to developing learners' awareness in relation to the language being learnt, the materials and methods being used, in short, the whole teaching process.

- Thirdly, learner training in self-assessment has to do with specific guidelines for evaluation. This again is very much a matter of judging the extent to which it is desirable to train learners in the use of a set of criteria defined by the teacher or some other authority, or whether it is individualization that one is trying to encourage. For example, if a group of learners are trained to monitor their oral performance with a view to improving fluency, group activities are audio-taped and analysed, and students work together to make comments and suggestions to each other. What happens, however, if one or more students in the group are particularly interested in perfecting their pronunciations of individual words or sounds? This seems to be a case in which the teacher would have to direct these learners to specific work in the area of pronunciation, and try to point out to them the aims of the fluency activities. In other words, the criteria being used for assessment would have to be very clear to students. At the

same time, the possibility of developing in different directions should be open to learners.

3.3- Innovation in Self-Assessment?

In view of the application of self-assessment as an innovative measure in the context of first year EFL learners in our university, it is necessary to analyse first what is meant by innovation.

In an article entitled "Innovation in language Testing", Alderson (1986) discusses the accountability of innovating. He suggests a number of questions which "innovators" should ask themselves, some of which we are going to use here to refer to the particular context of the introduction of self-assessment procedures.

1-"Are the changes that are claimed to be innovations actually new? In other words, do the self-assessment instruments used really promote learner awareness, development and responsibility for their learning process? Or are they the sort of tests which learners mark according to a key provided by the teacher and from which they get little but an overall score?

2-"Do the innovations, be the old or new actually work? If they work, do they work better than what they replace? The answer to this question implies, in the first place, the need for empirical observation. However, the question is not simply one of whether self-assessment works "better" than external exams in terms of measuring achievement or predicting performance. Because self-assessment is concerned with degree of motivation as well as with improving performance, evaluating the innovation becomes more difficult. This brings us to another of Alderson's questions.

3- "Might there be a need for innovation even if what is being replaced does work?" The answer seems to point to a "Yes!" if existing practices fulfil technical or institutional requirements but not the demands of educational objectives. In relation to testing, this might be the case if formal tests albeit being good measures of progress or achievement do not help to promote learner independence and involvement.

4-"Why do innovations come about? Who demands and causes them, and on what basis? "This has already been touched upon, in relation to the question of responsibility in evaluating learners' performance. This list of potential aspects to be considered is far from being a definitive one, given the complex nature of education and human relationships. It does certainly reflect, however, our own concerns and prejudices as well as aspects mentioned by different authors.

Alderson (1986:105) concludes his article entitled "Innovations in Language Testing? "by saying that" perhaps we should be looking for and fostering, not only innovations in language testing, but also and importantly, innovations through language testing." This was the line which this study tried to follow, starting from the premise that evaluation should not be seen as a necessary evil but as an intrinsic, constructive element of the teaching/ learning process.

4-Reservations as to the Use of Self Assessment: The Problem of Bias

When confronted with the question of whether students would be capable and therefore allowed to do self-evaluation, we believe most people — including a good number of teachers- would agree with Dickinson (1987:61) when he says that it is probably inevitable that most learners assessing themselves will be biased in their own favour, which will result in a certain distortion of test results.

This proposition, albeit plausible, does not seem to be always true, judging from Oskarsson's (1984: 32) report. According to him,

in at least four of the studies examined in detail the researchers found that the most proficient students tended to underestimate their ability and skill. Reported cases of overestimation tended to involve weak students to a greater extent than good ones.

Similar findings have been reported by McLeod (1983), who claims that good students tend to underestimate themselves because their awareness of language or high-level objectives give them the notion of what remains to be learned. Students who have experienced difficulties, on the other hand, tend to overestimate their ability since they cannot perceive a need for improvement. Underestimation of one's abilities may also indicate that students simply do not know or are not aware of what they can actually "do" with the target language. This is especially true of foreign language teaching in contexts where the target language is not commonly used outside the classroom, as it happens in our country.

When talking about bias in self- assessment, it is also important to distinguish, as Dickinson (1987:90) points out, between "the natural tendency of learners to interpret doubtful or ambiguous results in their own favour, and the deliberate falsification of results", or cheating. Dickinson goes on to add that cheating can only happen where there is a concern with showing results to someone else, in other words, if one is concentrating on products of instruction and on test scores.

III-Portfolio Assessment

1- What is a Portfolio?

Portfolio assessment is increasingly used in classrooms across the world. From its origins in the arts to its evolution through writing classes to foreign language classes, portfolio assessment is becoming nowadays a powerful tool for evaluating language skills.

A portfolio is a form of assessment that students do together with their teachers. According to Paulson and Meyer (1991:61), it is:

A purposeful collection of student work that exhibits the student's efforts, progress and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit and evidence of student self-reflection.

In this way, a portfolio is a living, growing collection of a student's work - each addition is carefully selected by the student for a specific reason which he/she will explain. It is also a selection of the student's work which shows growth over time. By reflecting on their own learning (self-assessment), students begin to identify the strengths and weaknesses in their work. These weaknesses then become improvement goals. In this sense Portfolios and instruction are closely linked; the assessment is integrated in the instructional process.

The criteria for selecting and assessing the portfolio contents must be clear to the teacher and the students at the outset of the process. The entries in an English Foreign Language (EFL) portfolio can demonstrate learning and growth in all language domains/skills, or can focus on a specific skill such as the ability to communicate genuinely in the target language. Portfolio assessment takes time and a sincere commitment from the teacher and student in order for the portfolio process to be successful.

2-Why Use Portfolio Assessment?

Introducing a new form of assessment requires a rationale acceptable to the educational institution, teachers, students and often parents. The overall purpose of the portfolio is to enable the student to demonstrate to others learning and progress. The greatest value of portfolios is that, in building them, students become active participants in the learning process and its assessment.

The following is a summary of some of the main points which support the use of assessment portfolios in the foreign language classroom.

- Portfolios are a form of alternative/authentic assessment, in which a student's progress is measured over a period of time in various language learning contexts. This allows a far more accurate form of language learning assessment than traditional testing which measures specific, isolated skills and abilities at a specific time, does not offer a full assessment of language learning. Portfolios can include evidence of specific skills and other items at one particular time and language performance and progress over time, under different conditions, in all four modalities (reading, writing, listening, and speaking).
- Portfolio assessment is incorporated fully into instruction, which has three educational benefits. First, linking assessment to instruction ensures that we are measuring what we have taught. Second, planning for and evaluating

portfolios reveals any weaknesses in instructional practices. For example, if the purpose of the portfolio is linked to making progress toward all areas of the National Standards, and, at the end of the marking period, there are no artifacts related to oral communication in the student portfolios, the teacher may decide to incorporate more oral communications work into the curriculum. Third, there is no time lost on assessment. Assessment is then a true learning experience, and not external to the learning process.

- Portfolio assessment promotes positive student involvement. As students create their portfolios, they are actively involved in their own learning. They reflect on their learning. This increased metacognition can have a positive impact on a student's self-confidence, facilitates use of learning strategies, and increases ability to assess and revise work. These factors tend to foster student motivation to continue studying and succeeding in language learning.

- Portfolios offer the teacher, student, and those who review the portfolios an in-depth knowledge of the student as a learner. This means that the teacher can individualize instruction for the student. Weak areas can be strengthened and areas of mastery built upon. Learners are involved in this process of tracking their learning and can take control of their learning.

- Portfolios can foster a relationship of mutual respect between teacher and student as they become partners in learning.

- Using portfolio assessment gives the teacher opportunities to involve parents in their children's language learning. Parental involvement is an important factor in educational success; providing language-related experiences and reviewing the portfolio are both ways in which parents can participate.

3-Stages in Implementing Portfolio Assessment

- **Identifying Teaching Goals to Assess through the Portfolio**

The very first and most important part of organizing portfolio assessment is to decide on the teaching goals. These goals will guide the selection and assessment of students' work for the portfolio. To do this, we have to ask ourselves "What do we want the students to learn?" and choose several goals to focus on; for example, general goals such as improvement in fluency of speech , and specific goals such as telling a story.

This stage is very important because teachers have to know what their goals are in terms of what the students will be able to do. Moreover, students have to know what they need to show evidence of in their portfolios.

It is even better if we do this fixing of goals together with the students, asking them, for example, what they need and want to achieve in the different language domains and skills. They will usually show good understanding of goals ("We need to understand the news." "We should be able to correct our pronunciation mistakes."), and hopefully these will then become common goals for teacher and class. Or, we can give a list of goals for the students to rank, and use the results for establishing the criteria for assessment.

- **Introducing the Idea of Portfolios to our Class**

We will need to present the idea of a portfolio to our class. We can start by explaining the word- from *portare* (carry) and *foglio* (sheet of paper). If possible, we ask a student of art, architecture or design to bring their portfolio; this will help convey the principle of a portfolio as a selection of a student's work, showing progress in different areas or skills.

It is also a good idea to show the students examples of English portfolios prepared by other classes, and, ideally, even a teacher's portfolio (showing, for example, the development of your work with the class).

It is worth directing students' attention, at this stage, to the main aspect of portfolios, which is their use as an assessment tool. We should try to ask our students how they feel about tests; whether they always feel the test truly represents what they know and can do with the language (they invariably bring up plenty of problems with traditional tests). Then, we should tell them we are going to assess them in a fairer way, which will show the many different skills, knowledge and ideas they have acquired.

We have to inform the students about how much weight the portfolio will have in their final grade and what it is going to replace (one or more of their tests, quizzes and/or projects).

We do not have to take on more than we can handle; we have to start with one class, and then expand when we feel ready. Students should be encouraged to put the right items into the portfolio; it is quality that counts, not quantity, and the main point of portfolio assessment is the thoughtful selection of evidence of learning.

- **Specifying Portfolio Content**

We have to specify what, and how much, has to be included in the portfolio - both core and options (it is important to include options as these enable self-expression and independence).

We also have to specify for each entry, written, audio and video-recorded items, artifacts, for example, an annotated drawing, a model, dialogue journals,

and how it will be assessed. The students should be acquainted with the scoring guides/rating scales that will be used before performing the task.

- **Giving Clear and Detailed Guidelines for Portfolio Presentation**

We have to explain the need for clear and attractive presentation and say how the portfolio will be graded. We need to remember that unfamiliar ways of teaching and assessment are potentially threatening and confusing to students. It is important to present the portfolio guidelines clearly, and to go over the guidelines periodically. Although all the guidelines - goals, content, timetable, etc should be presented to the class orally, so that they can discuss the procedure and ask questions, there should also be written guidelines to backup the points discussed and for reference while preparing the portfolio. It is helpful to prepare these guidelines in question-and-answer form.

- **Preparation Period**

Support and encouragement are required by both teacher and students at this stage. The students will get it from the understanding teacher. Teachers will get it by doing portfolio assessment as teamwork in their staff or joining or initiating a support group to discuss questions with colleagues as they arise.

We have to devote class-time to student-teacher meetings, to practising reflection and self-assessment and to portfolio preparation, since these may be new skills for most students, if not all the students.

Reflection and self-assessment do not come naturally to people who have had little practice in it, and require learner training. For example, the teacher has to encourage students to ask themselves: What did I learn from that activity? Which is my best piece? How can I improve this? This can be done by class brainstorming (what are some possible reasons for including an item in your

portfolio?) or in pairs - "portfolio partners" - who help each other select samples of their work (written comments on their work from a peer can also be included in the portfolio). Teachers should start with more structured forms of reflection and slowly proceed to more open reflective comments. This is training in a life-skill, and is well worth the time and effort spent in class.

The teacher should give guiding feedback. The finished portfolio may be due only at the end of the semester, but it is a good idea to set regular dates at which time several portfolio-ready items (i.e. with drafts and reflections) will be handed in, so that students know whether they are on the right track. Alternatively, you can have a portfolio project on a single unit of material so that both teacher and students will acquire experience in this kind of assessment over a shorter period of time.

To ensure that the portfolio represents the student's own work, some items can be done completely in class. You might also decide to have a test (preferably with corrected version) included as a core item together with reflection on what the student learned from doing the test and revising it. Furthermore, you may ask the students to explain in their reflections who helped them to improve their work (a peer, a parent, a spell-checker) and what they learned from revising their work.

- **Assessing the Portfolios and Giving Feedback**

Each portfolio entry needs to be assessed with reference to its specific goal(s). Since the goals and weighting of the various portfolio components have been clearly fixed in advance, in assessing the portfolios, a variety of assessment tools, such as rating scales and checklists for the different skills can be used. (see chapter 4, p 118) We can use these as they are, if they suit the fixed goals, or adapt them according to the new needs.

Self and peer-assessment can be used too as a tool for formative evaluation, with the students having to justify their grade with reference to the goals and to specific pages in the portfolio. This actually makes the teacher's job of assessing the portfolio much simpler because the student has done the groundwork of proving how far each goal is met in the portfolio. It takes some of the burden off the teacher and helps students to internalize criteria for quality work. Students can even generate their own report cards based on their portfolios.

After all the efforts that the students have invested in their portfolios, it is recommended that the teacher provides feedback on the portfolios that is more than just a grade. One possibility is to write a letter about the portfolio, which details strengths and weaknesses and generates a profile of a student's ability, which is then added to the portfolio. Another option is to prepare certificates which comment on the portfolio strengths and suggest future goals.

- **Student-Teacher Meetings**

An important element of the portfolio philosophy of shared and active assessment is that the teacher should have short individual meetings (tutoring sessions) with each student, in which progress is discussed and goals are set for a future meeting. Students and teachers should document these meetings and keep the goals in mind when choosing topics for future meetings. In this way, student-teacher conferences play an important role in the formative evaluation of a student's progress. They can also be used for summative evaluation purposes when the students present their final portfolios product and together with the teacher decide on a final grade. This is a student's chance to negotiate the portfolio grade using evidence of achievement according to the agreed goals. Notes from these meetings can be included in the portfolio as they contain joint decisions about the individual's strengths and weaknesses. These meeting

sessions can be prepared for in pairs where students practise presenting their portfolios.

4-Advantages and Disadvantages of Using Portfolio Assessment

Portfolio Assessment:

- allows the evaluators to see the student, group, or community as individual, each unique with its own characteristics, needs, and strengths;
- serves as a cross-section lens, providing a basis for future analysis and planning; by viewing the total pattern of the community or of individual participants, one can identify areas of strengths and weaknesses, and barriers to success;
- serves as a concrete vehicle for communication, providing ongoing communication or exchanges of information among those involved;
- promotes a shift in ownership, communities, and participants can take an active role in examining where they have been and where they want to go;
- offers the possibility of addressing shortcomings of traditional assessment and the possibility of assessing the more complex and important aspects of an area or topic;
- covers a broad scope of knowledge and information, from many different people who know the program or person in different contexts (for example, participants, parents, teachers or staff, peers, or community leaders).

However Portfolio Assessment:

- may be seen as less reliable or fair than more quantitative evaluations such as test scores;
- can be difficult or unfamiliar at first when we have to develop our own individualized criteria,

- can be just a miscellaneous collection of artifacts that do not show patterns of growth or achievement if goals and criteria are not clear;
- can be difficult to analyze or aggregate to show change, like any other form of qualitative data,
- can be very time consuming for teachers or program staff to organize and evaluate the contents, especially if portfolios have to be done in addition to traditional testing and grading.

Fear of time management issues is often the first reaction many teachers may have to the concept of portfolio use in the classroom, and with good reason: introducing anything new in the classroom takes time to plan and introduce to the learners. Portfolio assessment seems to require more time initially than other assessment instruments. However, one will probably find that the organization that we need to introduce with portfolio assessment will reduce on-going classroom planning and preparation because much of this work has been done in advance.

Further more we have to bear in mind that the time students spend developing their portfolios is not time lost to instruction and learning. It is through portfolio development that students use the skills and knowledge that are part of the curriculum. Additionally, while the class is working on portfolios, the teacher can offer mini-lessons to students in areas in which they need assistance; this maximizes teaching/learning time in class.

Changes in instruction can be difficult and time-consuming in the short term. In the long term, however, the benefits of portfolio use to teaching and learning are tangible and outweigh, in many teachers' minds, the initial challenges of implementing this new assessment tool.

In sum, this Alternative or New Vision to assessment can be exemplified as follows:

Old Vision

- Tests that test.
- One answer, one way correctness.
- Focus on language.
- Teacher-centered.
- Isolated skills.

New Vision

- Tests that also teach.
- Open-ended, multiple solutions.
- Focus on communication.
- Learner-centered.
- Integrated skills.

Conclusion

The reader may notice that different techniques for measuring oral proficiency have been appraised. Although their coverage is wide and interrelated, only their applicability in the language classroom has been considered. Through these techniques the learners will be stimulated to demonstrate their ability to present an impromptu talk, to converse, to discuss, to verbally reproduce the substance of a story heard or read. Teachers can succeed to apply and adopt the discussed techniques in the classroom by advocating authenticity of materials, relevance of situations in which speaking is an appropriate activity, cultural sensitivity, and other factors; they can make the learning environment as conducive to expression and language acquisition as possible.

Special attention is also paid to the possible uses of self- assessment and Portfolio, not only as an alternative to formal tests but basically and above all as a tool for learner independence. Possible problem inherent to the nature of self-evaluation were outlined, as well as anticipated problems related to the resources required for the implementation of such innovation for example, timing and training of staff and students.

CHAPTER SIX

NEEDS ANALYSIS

Introduction	169
I- The Students' Questionnaire	170
1-Aim of the Questionnaire	170
2-Description of the Questionnaire	170
3-Analysis of the Results	173
II-The Teachers' Questionnaire	216
1-Aim of the Questionnaire	216
2-Description of the Questionnaire	216
3-Analysis of the Results	220
Conclusion	256

NEEDS ANALYSIS

Introduction

The starting point of any test design should be to develop an analysis of learners' needs to the specification of language learning objectives. In the testing literature, there is a strong emphasis on the importance of test needs and purpose, and it is held that no one solution can accommodate the wide variety of possible test scenarios. It is argued that appropriately differentiated tests in different skills areas need to be made available for evaluating different groups of learners with different target situation needs. To measure Oral proficiency adequately in each situation, account must now be taken of teachers' and learners' opinions. We also have to ensure that the sample of the proposed language test is as representative as possible for the population to whom it is designed.

Different methods might be followed to collect information, such as surveys, questionnaires, interviews, attitude scales, intelligence, tests, job analysis, content analysis, statistical analysis and fixing of objectives (Richterich and Chancerel 1987). The identification of learners' needs might be made by the learner, the teaching establishment, the user institution. This identification can be made before the course, but it is important to take into account that learners' needs may change while they are learning, because needs are not static, but rather dynamic. They should be taken as a continuous process in which discussion, participation and negotiation with learners is crucial (Breen and Candlin 1980), and as flexible procedure which responds to changes in the learners, the teachers and the circumstances in which they are working.

I-The Students' Questionnaire

1- Aim of the Questionnaire

We have devised a questionnaire in order to explore and **find out the** difficulties encountered by first year LMD students, Department of **English, Faculty of** Languages and Human Sciences, University of Bejaia, while their oral proficiency is being tested. This information will help **us eventually to** devise a suitable battery of tests for that population of students.

2-Description of the Questionnaire

The questionnaire was handed in to the 372 students involved in the first year L.M.D during a lecture session in an Amphitheatre under the supervision of their teacher and the researcher. Among the 352 handed back questionnaires, 251 were returned during the same day and 101 remaining copies were collected three days later.

The questionnaire consists of closed and open questions. Closed questions are those which allow the student to select a number of possibilities **offered** to him/her or answer by "Yes", or "No". Open questions are those which require a personal answer .The questionnaire is made up of six sections.

Section 1: General Questions (Q1-Q7): This section seeks information about how the students rate their oral proficiency in spoken English: five rating options were proposed ranging from: very good, good, average, poor or very poor (**Q1**); whether and why they like to be tested (**Q2 and Q3**); whether they feel nervous before the test and why (**Q4 and Q5**); whether they feel nervous during the test and if "Yes" why (**Q6 and Q7**).

Section 2: Principles of Testing(Q8-Q15): this covers questions which seek to find out whether and why testing is the only responsibility of the teacher (**Q8 and Q9**); whether they consider testing as checking their performance, diagnosing their weaknesses , teaching, or any other answer they were asked to specify(Q10); whether they fully exploit their oral testing sessions to enhance their oral performance and if "No", they have to explain **why(Q11 and Q12)**; whether they were regularly tested (Q13),whether they like being regularly tested and, if "Yes" they were asked to explain **why(Q14 and Q15)**.

Section 3: Basic Considerations in Testing (Q16---Q29): This section includes questions about whether the tests they are given are related to the topics covered in the course , if "Yes", whether it is always, sometimes or rarely, and if "No", why (**Q16 , Q17 and 18**), whether the results obtained in the tests reflect their true level and if "Yes" whether it is always, sometimes or rarely and, if "No" whether they are given another test, a remedial work or anything else, they have to specify(Q19, **20 and Q21**); and whether in the case another teacher had conducted the test, they would have performed better(Q22 **and Q23**),**whether** the test activities are set at an appropriate level of difficulty and if "Yes" , whether it is always, sometimes or rarely(Q24 **andQ25**); **whether the** test discriminates adequately between their performances at different levels of attainment and if "Yes" , whether it is always, sometimes or rarely(Q **26 and Q27**);whether the time allotted to the test is adequate ,if "No" what suggestions can be made(Q28 **and Q29**).

Section 4: The Marking Scheme(Q30-Q34): this section is composed of questions, seeking information about whether the students are informed about the marking scheme used by their teacher and if "Yes", whether it includes grammar, pronunciation, stress and intonation or fluency or any other aspect (**Q30 and Q31**), whether the marking scheme indicates clearly the marks

to be awarded for different parts of a question or the relative weighting of criteria that may be applicable (Q32); whether this marking scheme refers to their non linguistic features, such as attitude, gestures, way of dressing and if "Yes" , is it always, sometimes or rarely (Q33 and Q34).

Section 5: The Oral Tests (Q35-Q43): This section is made of questions which aim at finding out which testing techniques are used to test their oral proficiency: multiple choice questions, oral interview, group work ,role play or discussion or any other techniques they had to specify(Q35); whether the testing techniques used are adequate to test their oral proficiency(for each option they had to say "Yes" or "No") and if "No" , why (Q36 and Q37); whether they understand test directions in: . multiple choice questions, oral interview, group work ,role play and discussion for each option, they had to say "yes" or "no",if "No", (Q38 and Q39);whether they have heard of self Assessment ,if "Yes", whether they are able to evaluate their oral proficiency (Q40 and Q41);whether they are given the opportunity to evaluate their oral proficiency and ,if "Yes", is it through assessing their own home or class work, assessing their peers' home or class work (Q22 and Q43).

Section 6: Further Suggestions (Q44): Here, the students are required to give their recommendations and suggestions to improve the testing procedures.

3-Analysis of the Results

Section 01: General Questions

Question 01: How would you rate your oral proficiency in spoken English?

a-Very good

b-Good

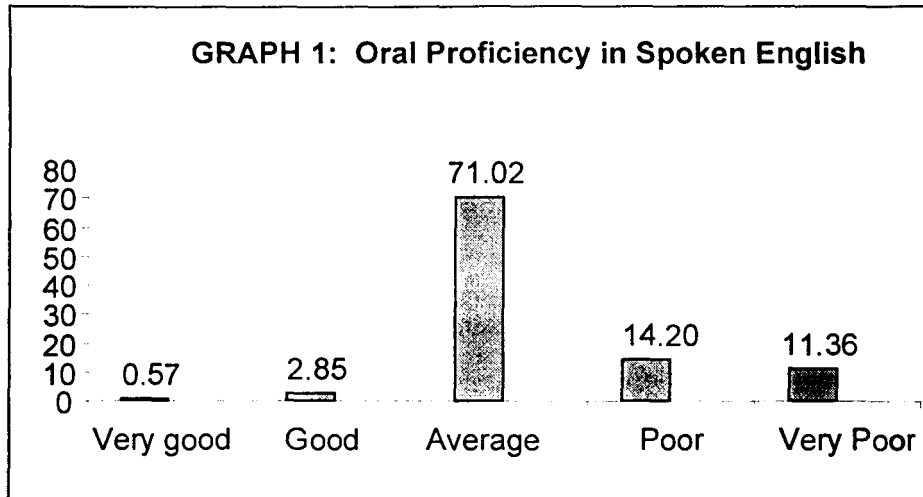
c-Average

d-Poor

e-Very poor

Table 06: Oral proficiency in Spoken English

Options	Number	%
Very good	2	0.57
Good	10	2.85
Average	250	71.02
Poor	50	14.20
Very Poor	40	11.36
Total	352	100



These results show that a very large majority (71, 02%) of the students think that their level in oral English is average. This may let us think that both teachers and learners need to work hard to reach better results.

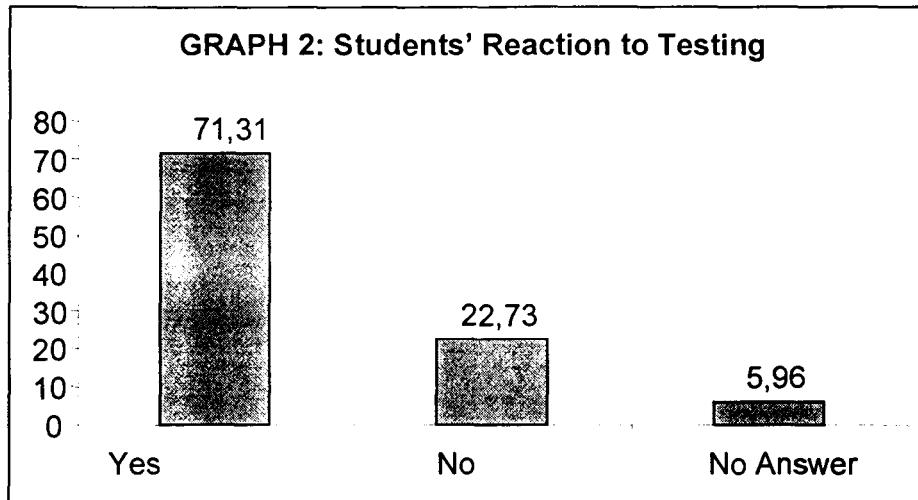
Question 02: Do you like to be tested?

Yes

No

Table 07: Students' Reaction to Testing

Options	Number	
Yes	251	71,31
No	80	22,73
No Answer	21	05,96
Total	352	100



The results indicate that 71, 31% of the questioned students declared that they like being tested. This shows that our students are really motivated to learn and are willing to be involved in testing situations.

Questions 03: Please, explain why

The 251 students who answered positively explained:

- we want to demonstrate what we know best in English (99 students)
- tests tell about our weakness and try to avoid them (136 students)
- tests teach us how to behave.(16 students) The

eighty students who answered negatively explained:

- the tests we are given are difficult(21 students)
- we are afraid to have bad marks (19 students)
- tests make us feel nervous (40 students)

Question 04: Do you feel nervous before the test?

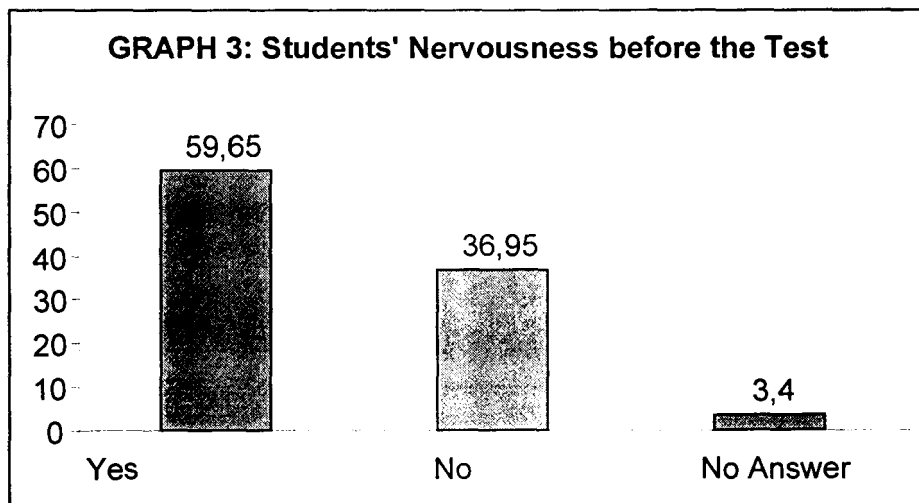
Yes

No

-Other: Please specify.

Table 08: Students' Nervousness before the Test

Options	Number	00
Yes	210	59,65
No	130	36,95
No Answer	12	3,40
Total	352	100



Nearly 60% of the students stated that they feel nervous before the test, which most probably implies that this will have an impact on the performance of the learners. We notice that 36, 93 % of them do not live a state of nervousness

before the test is administered. This may suggest that they feel confident in themselves in taking the test.

Question 05: If "Yes", please explain.

The justifications of the nervousness provided by the 210 students before taking the test are:

- we are not prepared for the test (120 students).
- we are afraid of the test; we are frightened by bad marks and failure (90 students).

Through the arguments provided by the respondents to question 5, we believe that there is a need to create better testing conditions and tutoring sessions will probably be of great help.

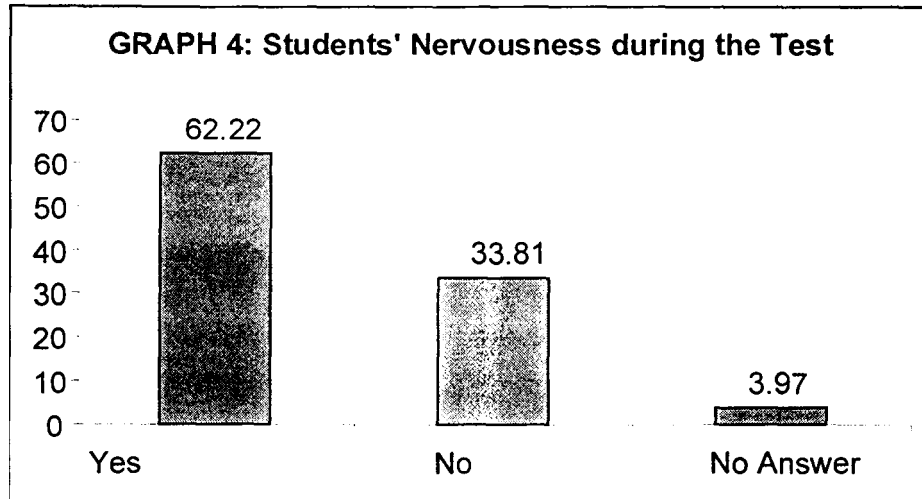
Question 06: Do you feel nervous during the test?

Yes

No

Table 09: Students' Nervousness during the Test

Options	Number	%
Yes	219	62,22
No	119	33,81
No Answer	14	03,97
Total	352	100



The results show that 62.21% of the students admitted being nervous during the test, which is slightly higher than the percentage of students who said they were nervous before the test (59.66%). This increase may be due to the nature of the test the learners are given.

Question 07: If "Yes", please, explain.

The 219 learners provided us with the following answers to justify their nervousness during the test:

- The questions are not very clear,(52 students)
- the test is too long and more difficult than we were expecting,(50 students)
- we are not prepared and trained for such tests,(42 students)
- We see some students cheating and this situation makes us very nervous, (40 students)
- some invigilators disturb and distract us during the test(35 students).

Section 2: Principles of Testing

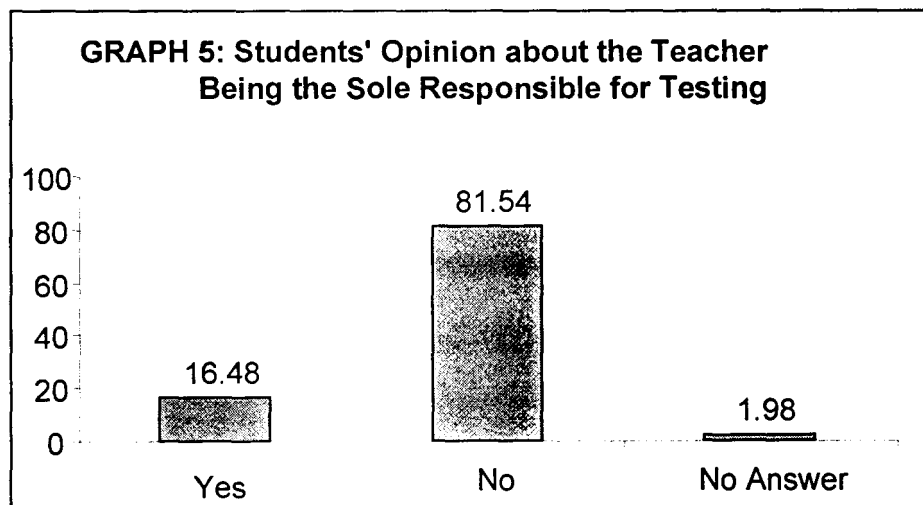
Question 08: Is the teacher the sole responsible of testing?

Yes

No

Table 10: Students' Opinion about the Teacher being the Sole Responsible for Testing

Options	Number	%
Yes	58	16.48
No	287	81.54
No Answer	07	01.98
Total	352	100



A very large number of the students (81, 53%) admitted **that testing is not the teachers'** only responsibility, and we consider this opinion as a positive argument if we want to reconsider and redefine the new expected roles of the

teachers and the learners in the process of assessment. Only a small portion of the informants i.e. 16, 47 % answered positively.

Question 09: Please, explain why

A hundred students did not supply any answer for this question. The rest of the learners provided us with the following arguments:

- We must take into consideration our opinions in any testing situation, (115 students).
- Testing our knowledge is also our responsibility since it is part of our learning process. (72 students)
- The 58 students who answered "Yes" said that their teachers can have a better judgment of their work and performance.

Question 10: -Do you consider testing as:

a-Checking your performance?

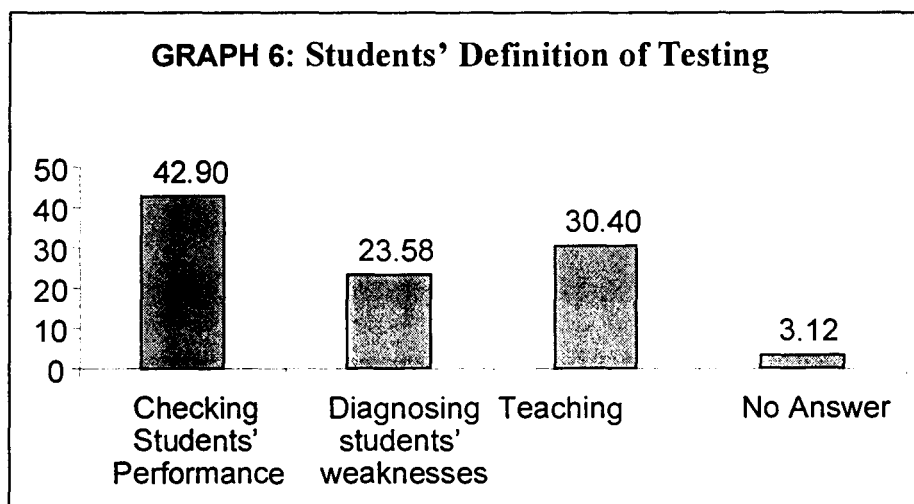
b-Diagnosing your weaknesses

c-Teaching?

d-Other: Please, specify.

Table 11: Students' Definition of Testing

Definition of Testing	Number	%
Checking students' performance	151	42,90
Diagnosing students' weaknesses	83	23,58
Teaching	107	30,40
No Answer	11	03,12
Total	352	100



42.90 % of LMD students think that testing is "Checking student's performance", 30.39 % of them mentioned "Teaching" and 23.57 % "Diagnosing students' weaknesses". Besides the proposed options, (5 students) said that testing the students is also finding out whether the teachers are doing their job properly or not.

Question 11: Do you fully exploit your testing sessions to enhance your oral performance?

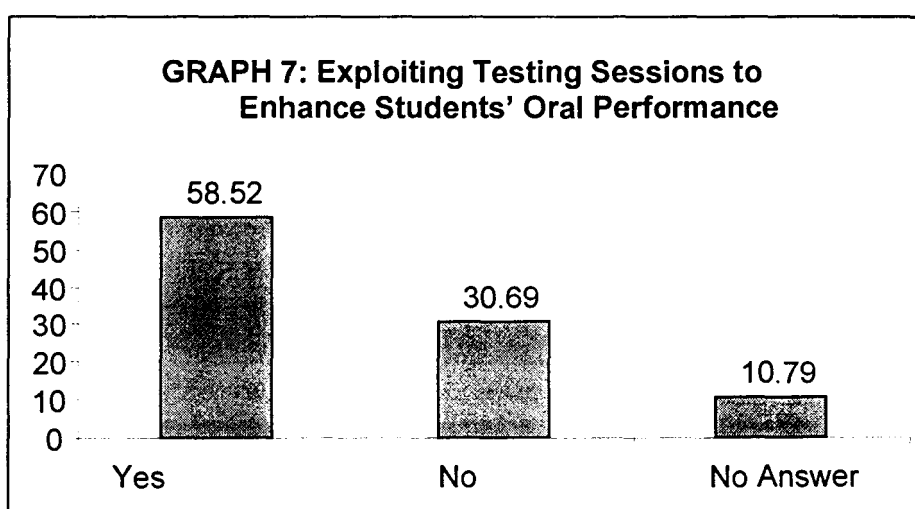
Yes

No

Table 12: Exploiting Testing Sessions to Enhance Students'

Oral Performance

Options	Number	00
Yes	206	58.52
No	108	30.69
No Answer	38	10.79
Total	352	100



Nearly sixty percent of the students said that they exploit fully their testing sessions to enhance their oral proficiency; this may be an indication of the students' interest and motivation to communicate in the target language.

Question 12: If "No", please, explain why

The following arguments were provided by the informants to justify their answer:

- the teacher dominates the speaking time,(68 students)
- It is always the same students who are given the opportunity to express themselves, (20 students)
- We are afraid and shy, (15 students)

- Sometimes we are intimidated by the teacher. (5 students)

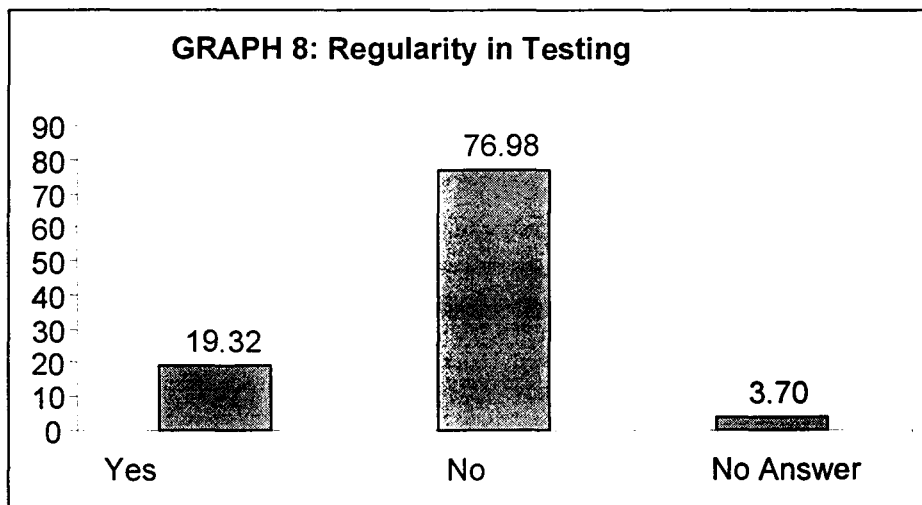
Question 13: Are you regularly tested?

Yes

No

Table 13: Regularity in Testing

Options	Number	%
Yes	68	19.32
No	271	76.98
No Answer	13	03.70
Total	352	100



The results to question 13 show that a large majority of the questioned learners are not regularly tested i.e. 76, 98%, and only 19, 31% answered positively. Our opinion is that new regulations of continuous assessment need to be fixed, so that the students' knowledge can be regularly **evaluated**.

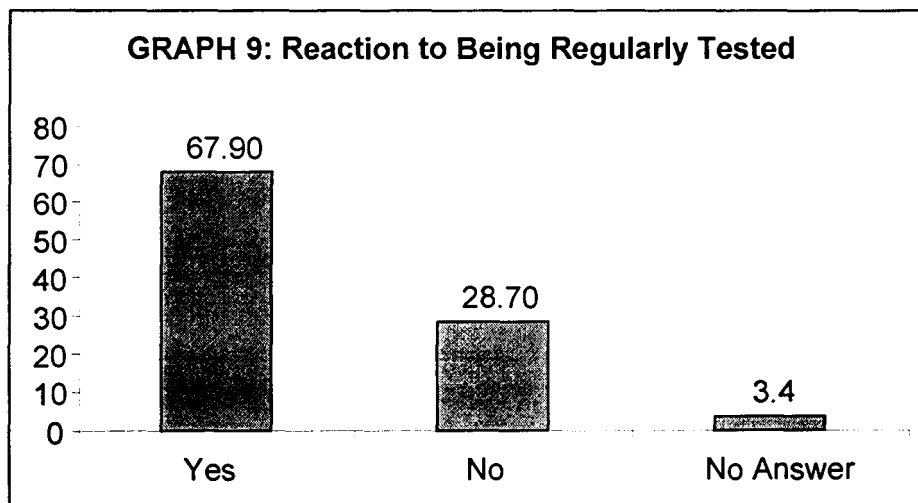
Question 14: Do you like being regularly tested?

Yes

No

Table 14: Reaction to Being Regularly Tested

Options	Number	%
Yes	239	67.90
No	101	28.70
No Answer	12	03.40
Total	352	100



The results give a clear indication about the desire of the first year LMD students to be regularly tested 67.90% answered positively. Such attitude may reflect the students' motivation to learning.

Question 15: If "Yes", please, explain why

The arguments evoked by the students who are willing to be tested regularly are:

- We want to find out our difficulties and weaknesses,(98 students)
- We want to perform and know our progress, (89 students)
- It is during oral session that we have the opportunity to put into practice what we know.(32 students)

Section 3: Basic Considerations in Testing

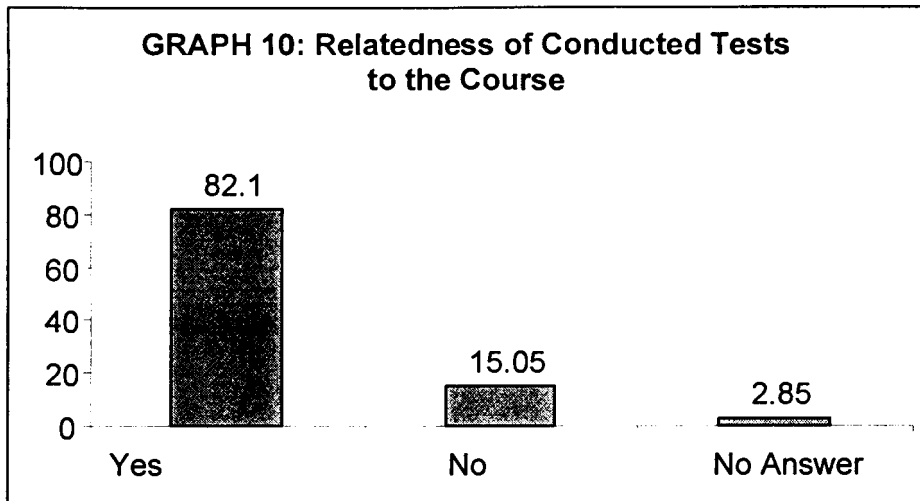
Question 16: Are the tests you are given related to the topics covered in the course

Yes

No

Table 15: Relatedness of Conducted Tests to the Course

Options	Number	
Yes	289	82.10
No	53	15.05
No Answer	10	02.85
Total	352	100



The indication that we can have from the answer to question 16 is that 82, 10% of the students stated that the tests they are given are related to the topics covered in the course, this percentage correlates with the teachers' answer for the same question.(see the teachers' questionnaire, Question 08)

Question 17: If "Yes", is it:

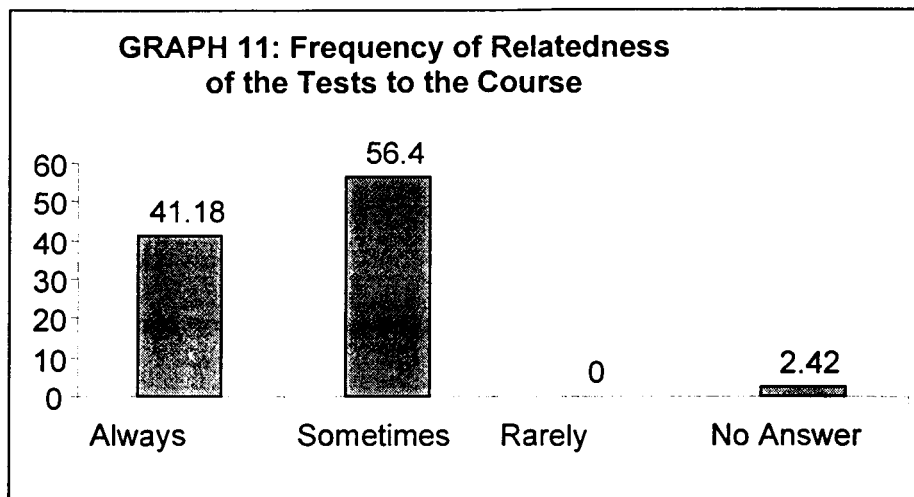
a -Always

b - Sometimes

c- Rarely

Table 16: Freauencv of Relatedness of the Tests to the Course

Frequency	Number	%
Always	119	41.18
Sometimes	163	56.40
R a r e l y		
No Answer	07	02.42
Total	289	100



Among the 289 students who stated that their tests are related to the topics covered in the course, 41, 17% of them said that this happens always. We think that the teachers of oral expression need to deal more and more with topics which have already been introduced to the students; this will probably give the learners more chance to communicate in the target language.

Question 18: If "No", Please, explain why

19 learners of those who answered "No" this question did not give any explanation to their answer. The most repeated answers that we collected were:

- we do not know ,(16 students)
- ask the teacher who has the answer. (18 students)

From the information provided to question 18, we can conclude that there is a need for the population under study to establish a positive learning environment based on communication.

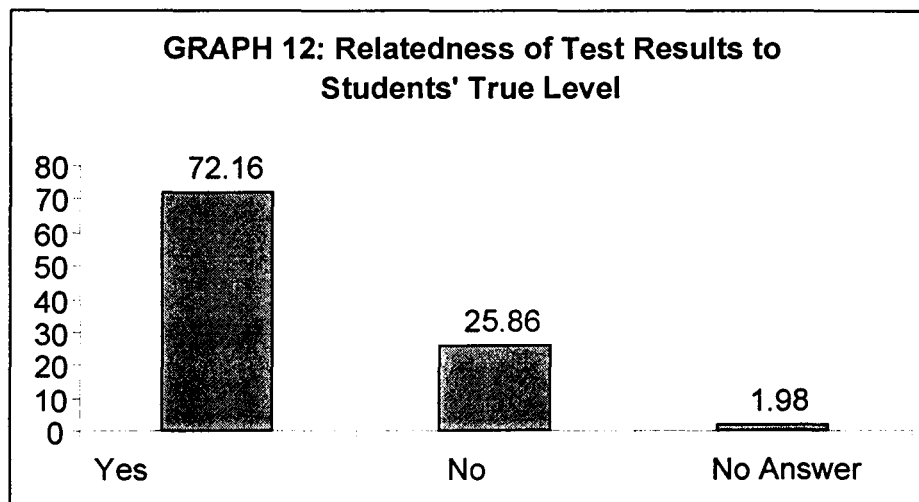
Question 19: Do the results obtained in the tests reflect your true level

Yes

No

Table 17: Relatedness of Test Results to Students' True Level

Options	Number	
Yes	254	72.16
No	91	25.86
No Answer	07	01.98
Total	352	100



A large majority of the students 72, 15% said that their test results reflect their true level .This may indicate that the questioned learners are aware of their strengths and weaknesses. The results may also stand for the fairness of the teachers' testing system.

Question 20: If “Yes”, is it:

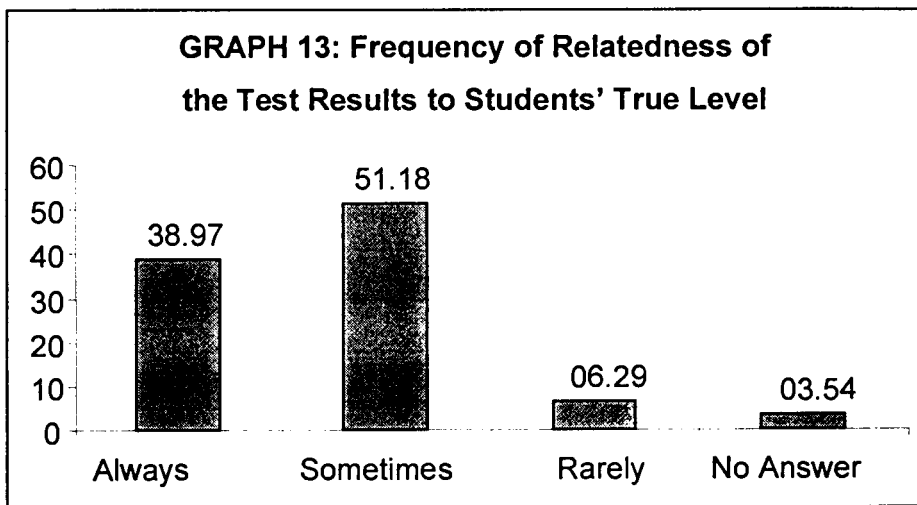
a -Always

b - Sometimes

c- Rarely

Table 18: Frequency of Relatedness of the Results to Students' True Level

Frequency	Number	%
Always	99	38.98
Sometimes	130	51.18
Rarely	16	06.30
No Answer	09	03.54
Total	254	100



When asked about the frequency of the relatedness of the test results with their true level, 38, 97 % of the students said that this happens always. We believe that this rate is less than average, so, teachers need to reconsider their way of testing; mostly when we notice that 51, 18 % responded sometimes.

Question 21: If "No", are you given:

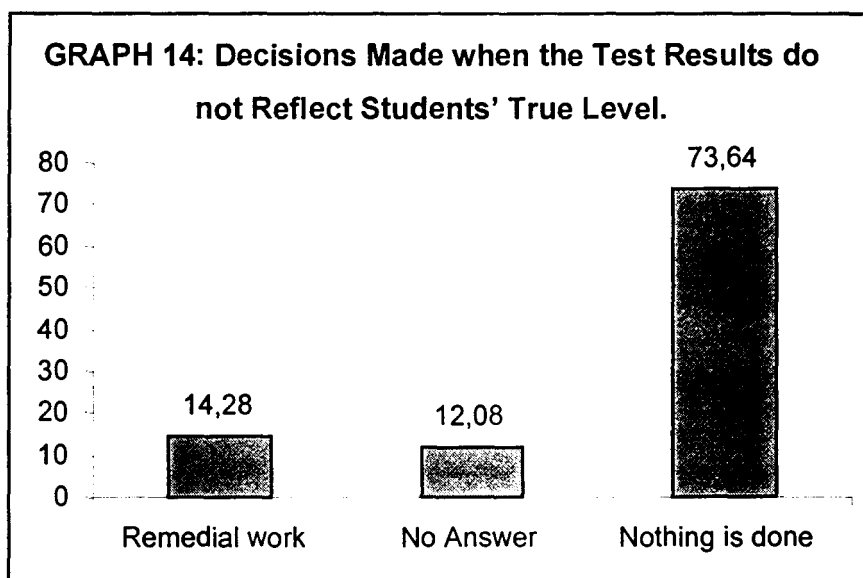
a-Another test

b-Remedial work

c-Other: Please, specify:

Table 19: Decisions Made when the Test Results do not Reflect Students' True Level_

Options	Number	
Another Test		
Remedial work	13	14.28
No Answer	11	12.08
Nothing is done	67	73.64
Total	91	100



Among the 91 students who answered "No" to the question 19, we have been surprised to find out that 67 students (73.64) specified that nothing is done for them when they obtain bad results, they added saying that they are rather neglected. This percentage indicates that there is an urgent need to devote more attention and care to learners who face difficulties in the process of learning.

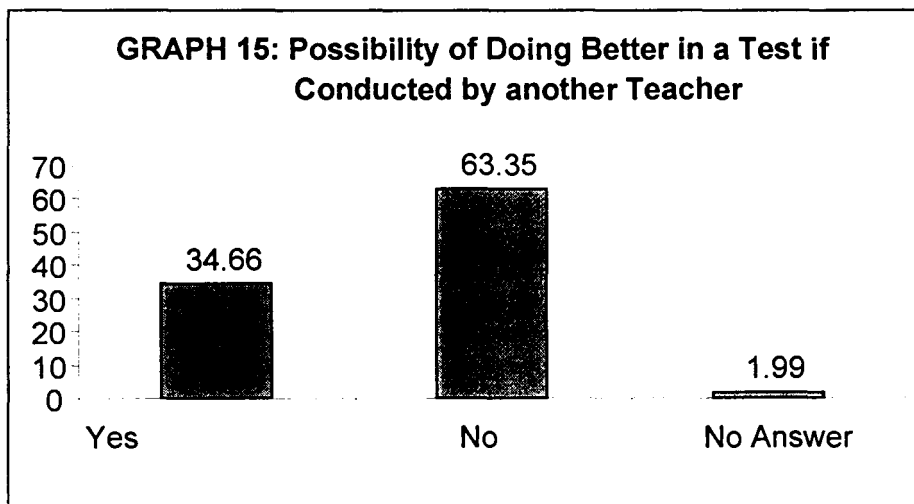
Question 22: If another teacher had conducted the test, would you have done better?

Yes

No

Table 20: Possibility of Doing Better in a Test if Conducted by another Teacher

Options	Number	%
Yes	122	34.66
No	223	63.35
No Answer	07	01.99
Total	352	100



We can notice that more than 63% of the students do not believe that if they are assessed by another teacher they will obtain a better mark. These results may tell us that the learners believe in their teacher and in the way he/she is assessing their knowledge in the target language.

Question 23: Please, explain why.

When we asked the students to justify their answers for those who answered "No" to question 22, we have obtained the following answers;

- Our teacher is the only one who can judge us because she knows the things we have dealt with in the classroom, (203 students)
- We do not know, simply because we have not tried such testing Situation, (20 students)
- The 122 students who answered positively said that they wish to be assessed by anybody else because their teacher is too subjective.

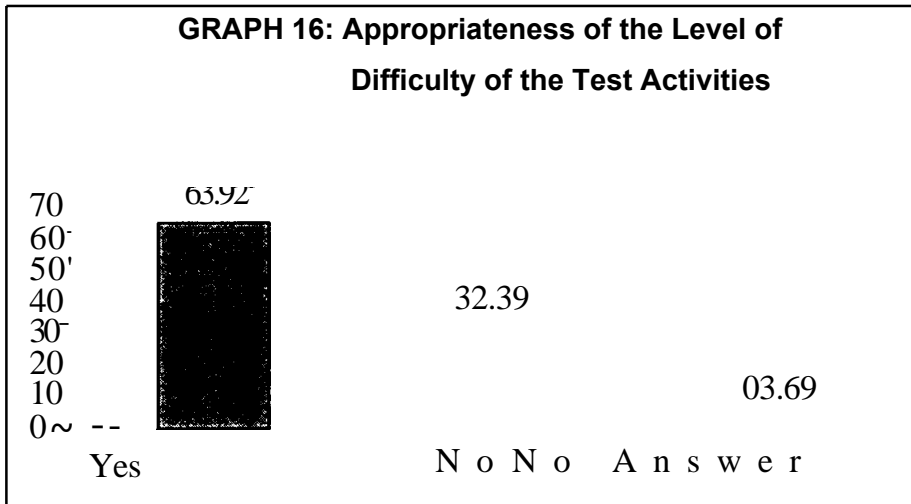
Question 24: Are the test activities set at an appropriate level of difficulty?

Yes

No

Table 21: Appropriateness of the Level of Difficulty of the Test Activities

Options	Number	
Yes	225	63.92
No	114	32.39
No Answer	13	03.69
Total	352	100



More than 63% of the students who answered this question recognized that the test activities set are at the appropriate level of difficulty. These results correlate with the teachers' answer. (Question 23 page 235)

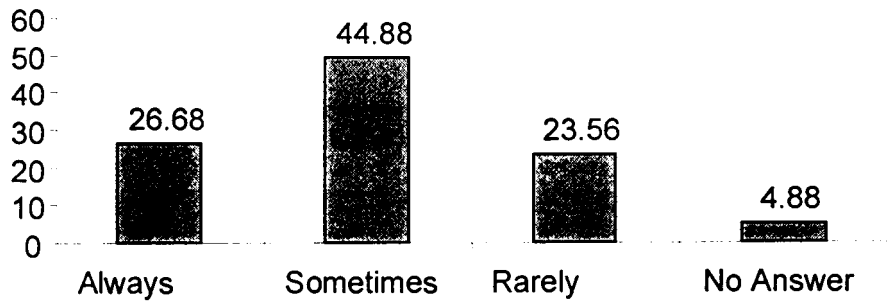
Question 25: If "Yes", is it:

- a -Always
- b - Sometimes
- c- Rarely

Table 22: Frequency of Appropriateness of the Level of the Test Activities

Frequency	Number	%
Always	60	26.68
Sometimes	101	44.88
Rarely	53	23,56
No Answer	11	4.88
Total	225	100

GRAPH 17: Frequency of Appropriateness of the Difficulty of the Test Activities



Through the obtained results , we can notice that only 26, 66% of the students who answered "Yes" to question 25 said "always"; we believe that this is a very low percentage mostly if one has the ambitious vision to deny and forget some of the traditional procedures of language measurement which do not give the pedagogical expected results.

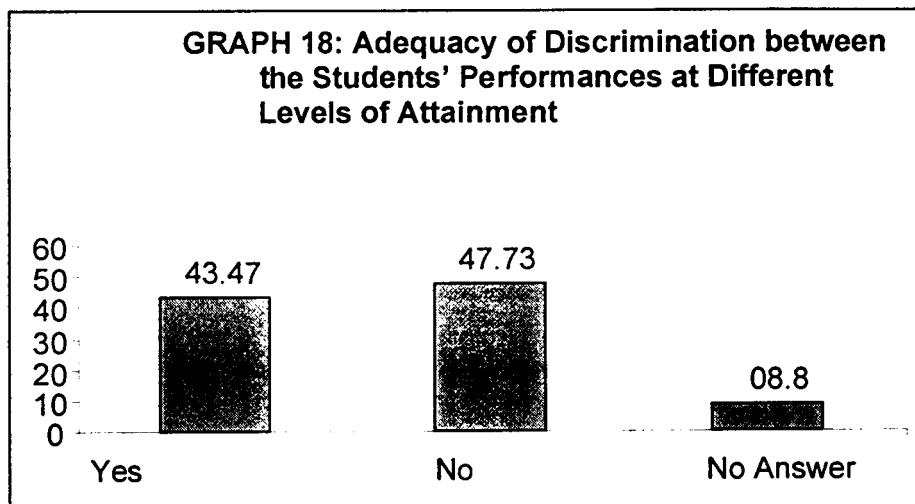
Question 26: Does the test discriminate adequately between your performances at different levels of attainment?

Yes

No

Table 23: Adequacy of Discrimination between the Students' Performances at Different Levels of Attainment

Options	Number	%
Yes	153	43.47
No	168	47.73
No Answer	31	08.80
Total	352	100



The results show us that 47, 73% of the students told us that the test does not adequately discriminate between their performances at different levels of attainment; this may let us think that the way tests are devised for this population of students must be reconsidered. Only 43, 46% gave a positive answer.

Question 27: If "Yes", is it:

a -Always

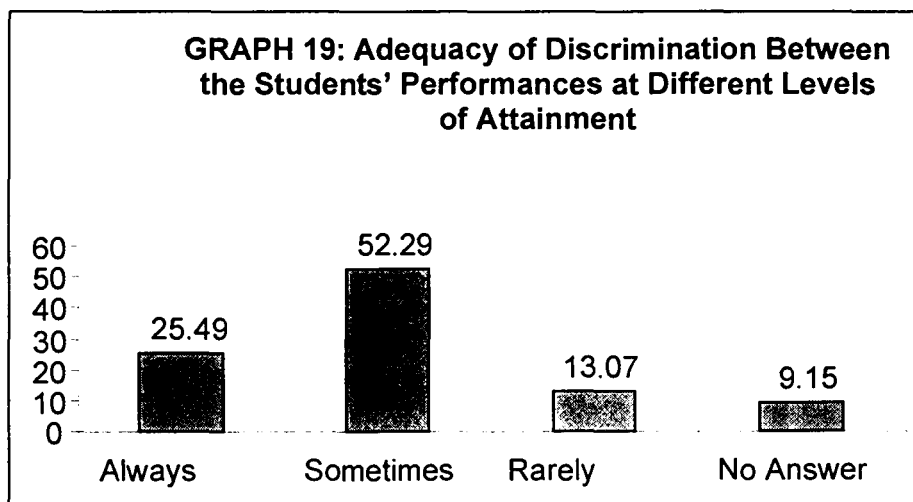
b - Sometimes

c- Rarely

Table 24: Frequency of Adequacy of Discrimination between the Students'

Pe. rfnrmaw.nnc. n é Tl: fen rnni I "nun] o ni Afin;nrnnn±

Frequency	Number	%
Always	39	25.49
Sometimes	80	52.29
Rarely	20	13.07
No Answer	14	09.15
Total	153	100



We found out that among the students who answered "Yes" question 26, only 25, 49 % answered always, whereas more than 52% have opted for the option "sometimes". This may mean that the devised tests lack the elements of reliability and validity.

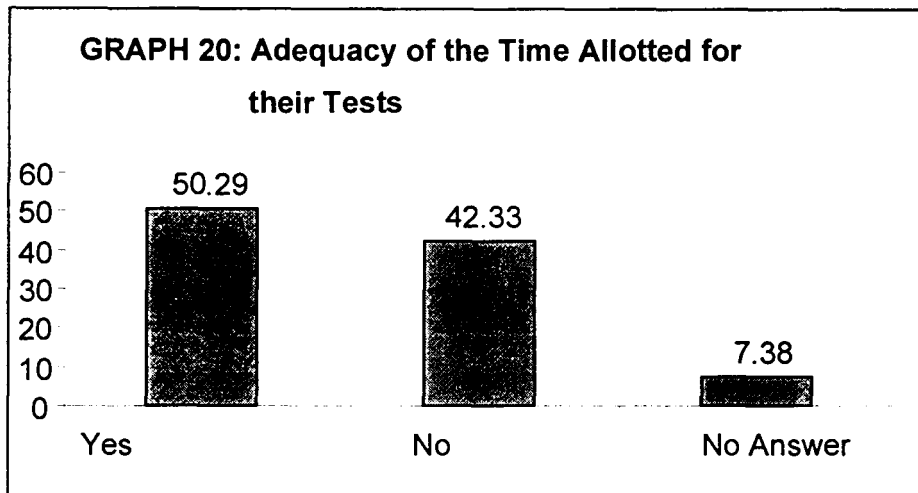
Question 28: Is the time allotted for the test adequate?

Yes

No

Table 25: Adequacy of the Time Allotted

Options	Number	%
Yes	177	50.29
No	149	42.33
No Answer	26	07.38
Total	352	100



50, 85% of the informants have stated that the time allocated to the test is adequate. But at the same time 42.33% answered negatively. However, we think that it is necessary that teachers involved in this study reconsider the element of time allocation when planning their tests.

Question 29: -If "No", what do you suggest?

The 75 students, who complained about the inadequacy of the time allotted for their tests, did not provide any answer for this question. The remaining 74 respondents were asking for more time to be devoted for the tests they are given.

Section 4: The Marking Scheme

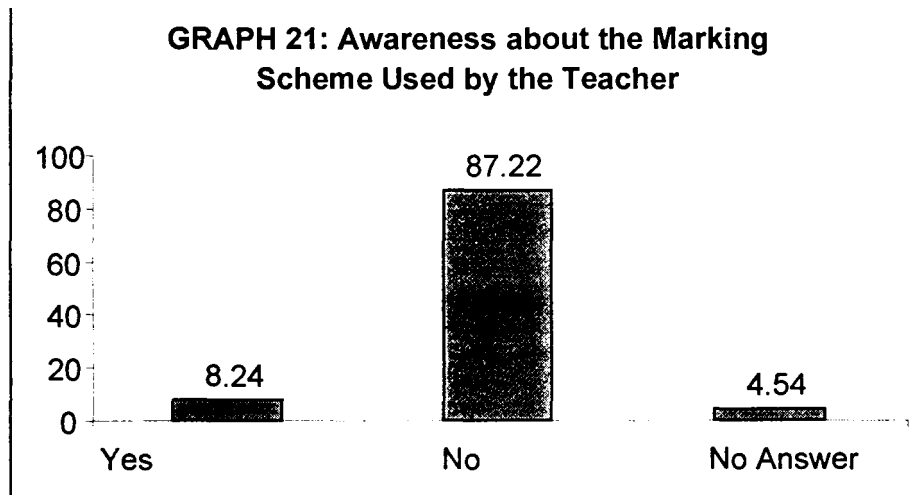
Question 30: Are you informed of the marking scheme used by your teacher?

Yes

No

Table 26: Awareness about the Marking Scheme Used by the Teacher

Options	Number	%
Yes	29	08.24
No	307	87.22
No Answer	16	04.54
Total	352	100



A large majority of the students i.e. more than 87% are not informed of the marking scheme used by their teacher, and only 08, 23% gave a positive answer. We believe that this manner of dealing with learners' measurement must be radically reconsidered if we are willing to obtain better results.

Question 31: If "Yes", does it include?

a- Meaning

b- Grammar

c- Pronunciation

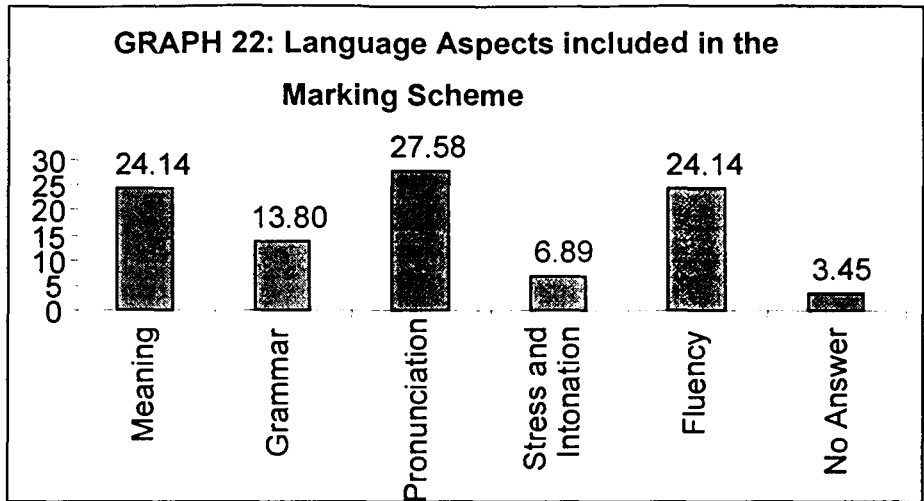
d- Stress and intonation

e- Fluency

f- Other: Please, specify :

Table 27: Language Aspects included in the Marking Scheme

Language aspects tested	Number	%
Meaning	07	24,14
Grammar	04	13,80
Pronunciation	08	27,58
Stress and intonation	02	06,89
Fluency	07	24,14
No Answer	01	03,45
Total	29	100



We note that all the language aspects are included in the marking scheme. But we notice that more importance is given to pronunciation (27.58%) and meaning and fluency. (24, 14% each)

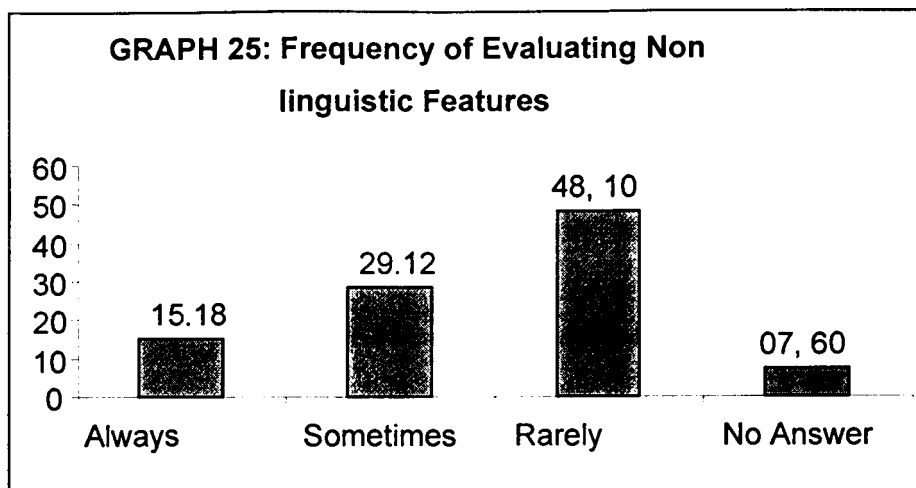
Question 32: Does the marking scheme indicate clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable?

Yes

No

Table 28: Clarity of Indication of the Marking Scheme in Relation to the Marks to be Awarded for Different Parts of a Question or the Relative Weighting of Criteria that Might be Applicable.

Options	Number	%
Yes	59	16.76
No	277	78.70
No Answer	16	04.54
Total	352	100



Among the 79 students who mentioned that their teacher considers their non linguistic features while being tested, 48, 10% said rarely. This let us think that teachers involved in this study ignore linguistic feature while testing their students.

Section 05: The Oral Test

Question 35: Which testing techniques are used to test your oral proficiency?

a-Multiple Choice Questions

b-Oral Interview

c- Group work

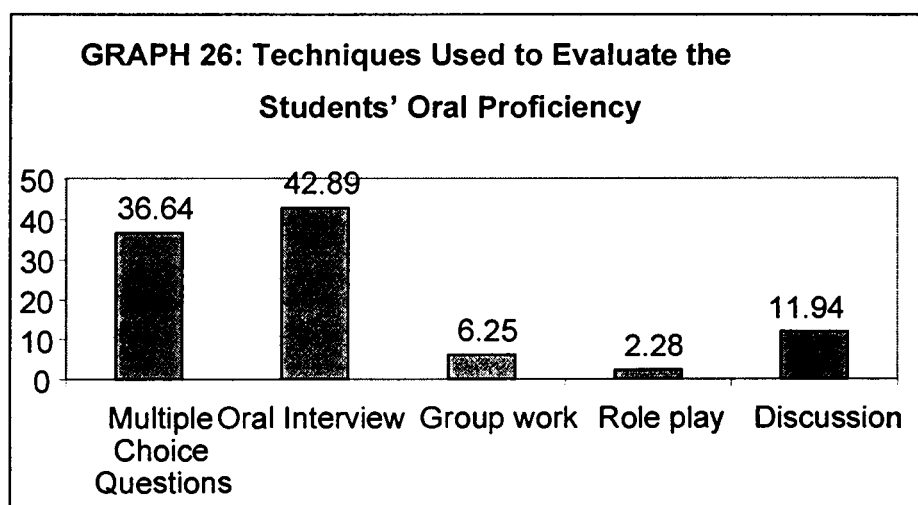
d-Role play

e-Discussion

f-Other: Please, specify

Table 31: Techniques Used to Evaluate the Students' Oral Proficiency.

Techniques	Number	%
Multiple Choice Questions	129	36.64
Oral Interview	151	42.89
Group work	22	06.25
Role play	8	02.28
Discussion	42	11.94
Total	352	100



We can notice that the most used techniques to evaluate students' oral proficiency are: Oral interview 42, 89 %, and MC Q we have 36, 64%. This indicates clearly that there is no innovation in the way the teachers of the Department of English evaluate their students. We noted that the respondents did not make any combination among the proposed techniques and they opted for one technique at a time.

Question 36: Are these testing techniques adequate to test your oral proficiency?

a- Multiple Choice Questions Yes

No

b- Oral Interview Yes

No

c- Group work Yes

No

d-Role play Yes

No

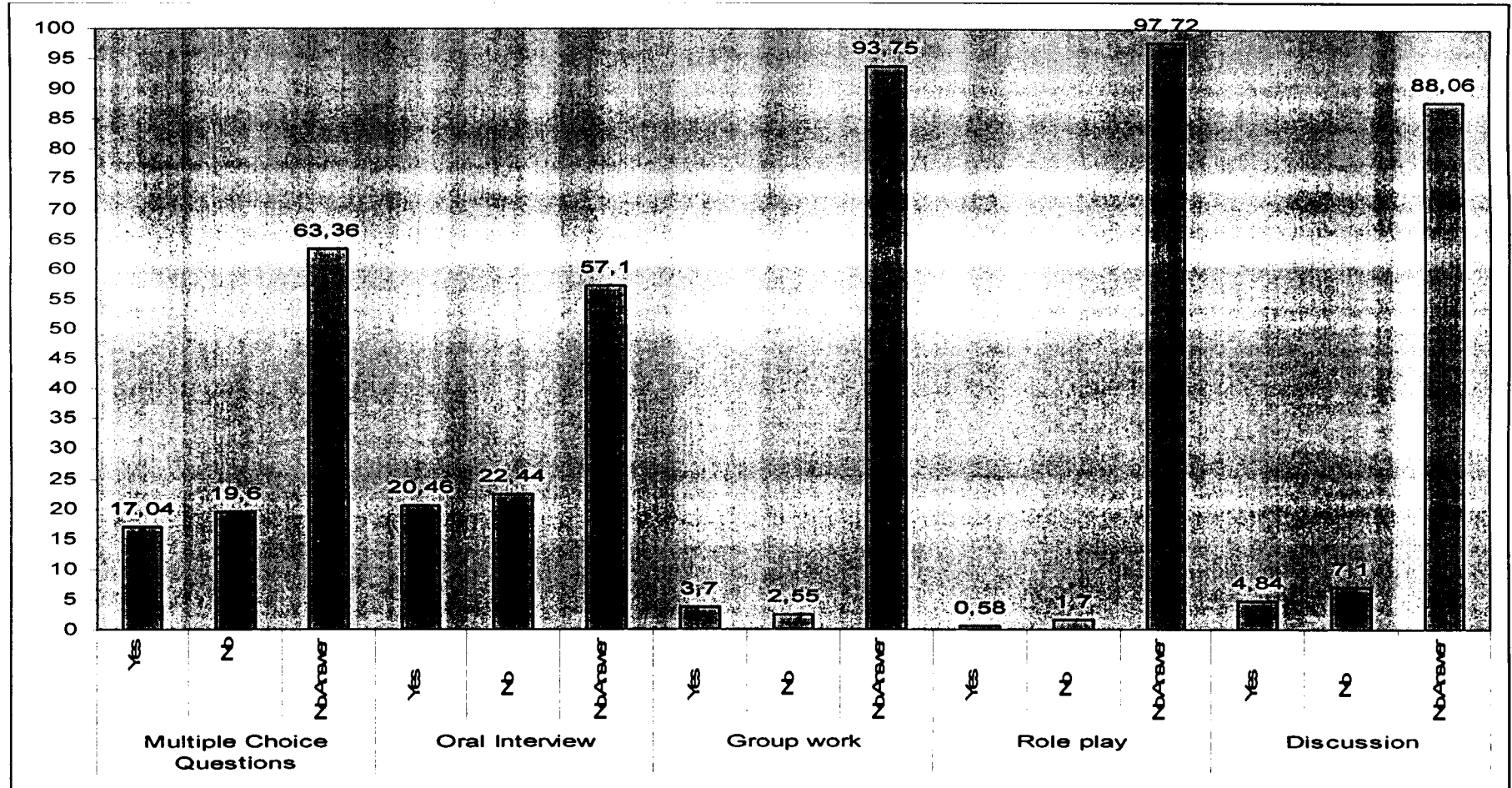
e- Discussion Yes

No

Table 32: Adequacy of the Techniques used to Evaluate the Students' Oral Proficiency

Techniques	Options	Number	%
Multiple Choice Questions	Yes	60	17.04
	No	69	19.60
	No Answer	223	63.36
	Total	352	100
Oral Interview	Yes	72	20.46
	No	79	22,44
	No Answer	201	57.10
	Total	352	100
Group work	Yes	13	3.70
	No	09	2.55
	No Answer	330	93.75
	Total	352	100
Role play	Yes	02	0.58
	No	06	1.70
	No Answer	344	97.72
	Total	352	100
Discussion	Yes	17	4.84
	No	25	7.10
	No Answer	310	88.06
	Total	352	100

GRAPH 27: Adequacy of the Techniques used to Evaluate the Students' Oral Proficiency



Through the table above which represents the adequacy of the techniques used to evaluate students' oral proficiency, one can notice that there is mostly a shared opinion about those who are satisfied and those who are not. If we take for example the 151 students who referred to the technique of Oral Interview, we will realize that 47, 68% answered "Yes" and 52, 31% "No".

Question 37: If "No", please, explain why.

Nine students mentioned that they do not appreciate group work because some elements of the group are lazy and are not cooperative. 65 unsatisfied learners said that Multiple Choice Question and True or False questions are not adequate but rather boring because of their guessing nature.

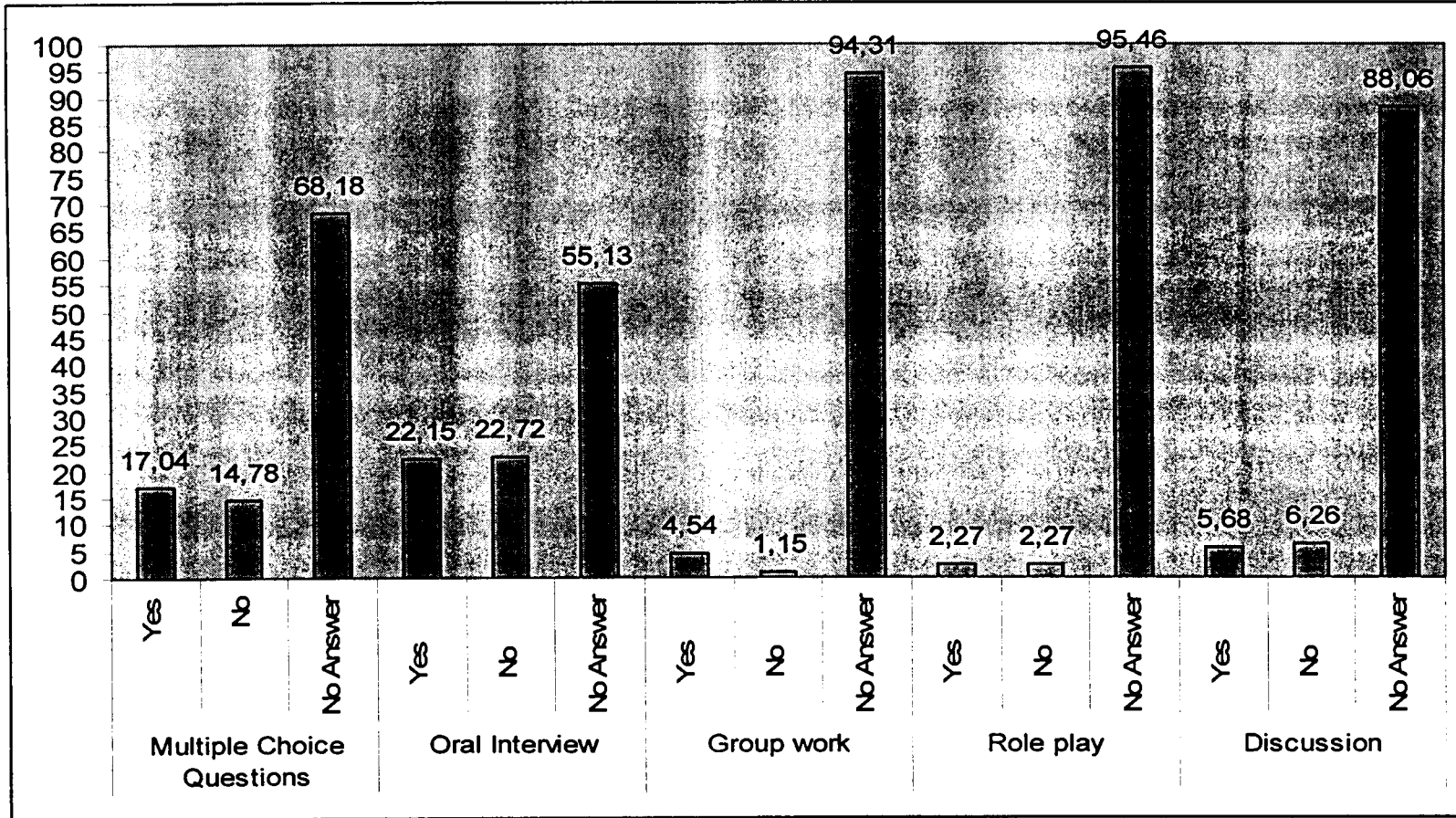
Question 38: -Do you understand the test directions in:

a- Multiple Choice Questions	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
b- Oral Interview	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
c- Group work	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
d-Role play	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
e- Discussion	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>

Table 33: Students' Understanding the Test Directions

Techniques	Options	Number	%
Multiple Choice Questions	Yes	60	17.04
	No	52	14.78
	No Answer	240	68.18
	Total	352	100
Oral Interview	Yes	78	22.15
	No	80	22.72
	No Answer	194	55.13
	Total	352	100
Group work	Yes	16	4.54
	No	04	1.15
	No Answer	332	94.31
	Total	352	100
Role play	Yes	08	2.27
	No	08	2.27
	No Answer	336	95.46
	Total	352	100
Discussion	Yes	20	5.68
	No	22	6.26
	No Answer	310	88.06
	Total	352	100

GRAPH 28: Students' Understanding of the Test Directions



If we have a look to the data represented in graph 28, we will find out that there is nearly a balance between those who understand and those who do not understand the test directions. We believe that the involved teachers can be required to reconsider the way they devise their tests.

Question 39: If your answer is "No", please explain why.

Half of the informants for each technique did not answer this question, as for the other half they declared that the questions are very often ambiguous; they do not give clear indication of what the examiner is asking. Very often we are trapped; consequently we take the question to mean something different.

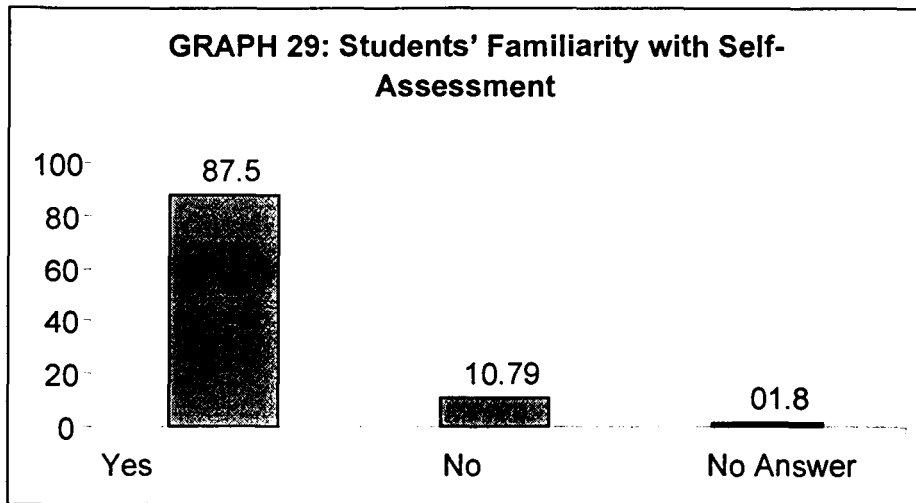
Question40: Have you heard of self —assessment?

Yes

No

Table 34: Students' Familiarity with Self -Assessment

Options	Number	%
Yes	308	87.5
No	38	10.79
No Answer	06	01.8
Total	352	100



We can note from the answers to question 40 that a very large number of the first year LMD students (87.50%) have already heard of self assessment. Here, we may need to mention that before starting the investigation, we gathered an amount of data about this type of assessment which we have handed to all the students concerned by this study.

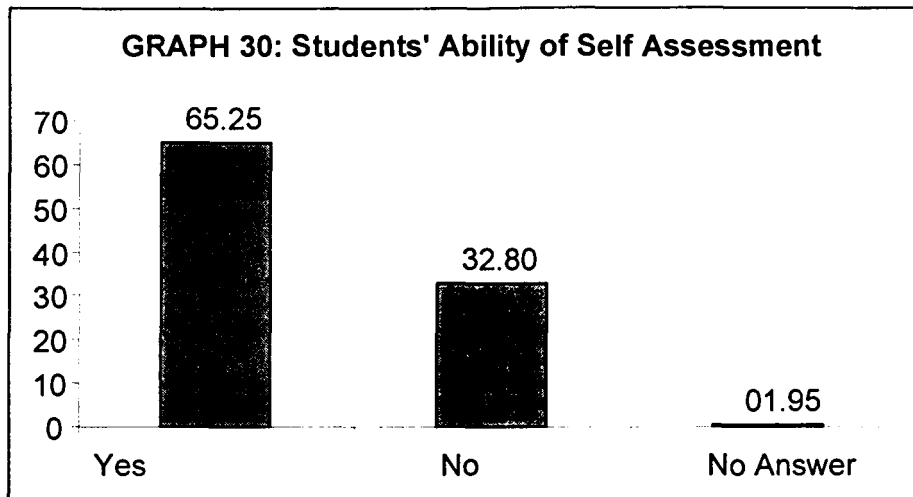
Question 41: If "Yes", are you able to evaluate your oral proficiency?

Yes

No

Table 35: Students' Ability of Self -Assessment

Options	Number	%
Yes	201	65.25
No	101	32.80
No Answer	06	01.95
Total	308	100



We have found out that 65.25% of the questioned students have declared to have the ability to measure their oral ability. But if we devote some of our pedagogical time to explaining to students the positive impact that this evaluation can have on their process of learning, more and more learners will be involved.

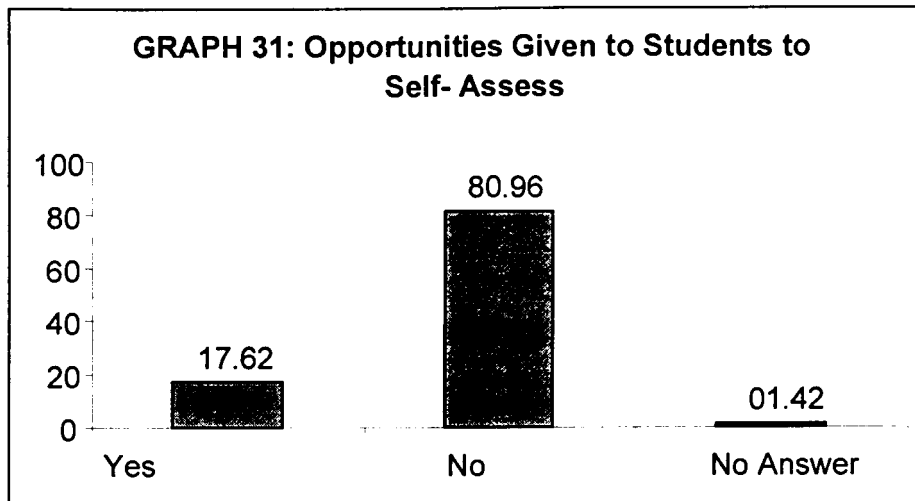
Question 42: Are you given the opportunity to evaluate your oral proficiency?

Yes

No

Table 36: Opportunities Given to Students to Self -Assess

Options	Number	%
Yes	62	17.62
No	285	80.96
No Answer	05	01.42
Total	352	100



We can realize through the answers provided that only 17.62% are given the opportunity to self assess their ability to communicate, which is undoubtedly a small percentage. This may indicate that some of the teachers of the first year LMD students ignore, or are not prepared for such system of evaluation.

Question 43: If "Yes" is it through:

a-Assessing your own home or class work b-

Assessing your peers' home or class work c-

Other, Please, specify

50% of the respondents (31 students) did not supply any answer to this question. The rest of the students who are offered the opportunity to self assess provided us with the following answers:

- We are asked to assess our own home and class work (18 students)
- We are asked to assess our peers' home and class work (13 students)

Section 4: Further Suggestions.

Question 44: Would you please add any suggestion or recommendation.

The propositions came as follows:

A large number of the learners 65% expressed their desire and interest to be introduced to the technique of self Assessment.

Nearly all questioned students (95 %) expressed the wish to have more trained and qualified teachers who will be able to provide them with the appropriate tools and support so as to improve their speaking ability.

Many students complained about the fact that some of their teachers give their lectures in other languages than the target language. (80%) of the students wished their lectures to be delivered only in English so that they become more familiar with the language.

Some of the students 30% suggested that the number of hours of the Oral English course should be increased.

III-Teachers' questionnaire.

1-Aim of the Questionnaire

It is our strong belief that in order to diagnose effectively our first year students' difficulties while their oral English ability is being tested and develop an effective battery of tests, it is necessary to consider the teachers' opinions about the most suitable and adequate testing techniques to adopt for such population of learners.

2-Description of the Questionnaire

The questionnaire includes forty six questions and it was given to 13 teachers of oral expression who teach in the department of English. This questionnaire consists of seven parts.

The Questionnaire was handed to the totality of the teachers of oral expression of the Department .Two days later; the thirteen copies of the Questionnaire were collected back.

Section 1: General Questions

This section includes four questions (**Q1 to Q4**): seeking information about the teaching experience in general (**Q1**); how long they have been teaching the oral expression. (**Q2**); how long they have been teaching oral expression in the first year. (**Q3**), and finally whether they have any special training to teach oral expression. (**Q4**).

Section 2: Principles of Testing

This part includes seven questions (**Q5 to Q12**): related to whether the teacher is the sole responsible of testing the teachers were asked to explain why (Q5 and Q6); whether they consider testing as: Checking student's performance, diagnosing students' weaknesses, or teaching. As they were asked to specify any other definition of the term testing (Q7); whether they exploit fully their testing sessions to enhance their students' oral performance and if "Yes», they have to explain how (**Q8 and Q9**); if "No", they were asked to say why (Q10), whether they regularly test their students, and if "Yes", they were asked to explain why (**Q11 and Q12**).

Section 3: The Testing Procedure

This section is composed of four questions (**Q13 to Q16**): the purpose is to be informed whether they write their own tests and if "No", we asked them if they are provided by another colleague, selected from a test book or from any other source. (**Q13 and Q14**); whether they administer and conduct oral examinations alone and the respondents were asked to explain why (**Q15 and Q16**).

Section 4: Basic Consideration in Testing

this section contains eleven questions (**Q17 to Q27**): the teachers were asked whether the tests they conduct are related to topics they covered in the course and if "Yes", is it:always,sometimes or rarely(**Q17 and Q18**),if "No", they have to say why(Q19);whether the results obtained from the delivered tests reflect their students' true level, if "Yes" ,is it :always, sometimes or rarely(Q20 and Q21);if "No" , do they give: another test, a remedial work or any other

action which they have to explain(Q22),here the teachers were asked whether the test activities are set at an appropriate level of difficulty, if "Yes" , is it:always, sometimes or rarely (Q23 and Q24),whether the tests discriminate adequately between the performances of candidates at different levels of attainment°, and if "Yes", is it:always,sometimes or rarely(Q25 **and** Q26),and finally whether the abilities being rewarded are those which the activities are designed to assess(Q27).

Section 5: The Marking Scheme

This part is made of nine questions (**Q28 to Q36**): seeking information about whether the marking scheme anticipates responses of a kind that candidates are likely to make(Q28); whether the marking scheme indicates clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable(Q29);whether the marking scheme allows for possible alternative answers(Q30);and to what extent the marking scheme , by specifying performance criteria, reduces as far as possible the element of subjective judgement that the examiner has to exercise in evaluating candidates' answers (**Q31**); whether the marking schemes can be easily interpreted by a number of different examiners in a way which will ensure that all mark to the same standard(Q32);which of the following language aspects they test i.e.meaning grammar, pronunciation, stress and intonation, fluency , or any other aspect which they have to specify. They were asked to mention the order of importance i.e. (1 for the most important one to 5 the least important one.) (Q33); whether they evaluate non linguistic features, such as attitude, gestures, way of dressing, if "Yes", is it: always, sometimes or rarely (**Q34 and Q35**), if "No", they were asked to explain why (Q36).

Section 6: Oral Tests

This section which is a total of nine questions (**Q37 to Q45**): is collecting information about which of the following testing techniques teachers use to evaluate their students' oral proficiency i.e. Multiple choice questions oral interview, group work, role play or discussion, if they use other techniques, they were asked to mention them (Q37); whether they have heard of self assessment , if "Yes" ,they were asked if their students are able to evaluate themselves (**Q38 and Q39**); whatever the answer they provide, they have to explain why(Q40); whether students are given opportunities to evaluate their oral proficiency , if " Yes" , is this self assessment done through: assessing their own home or class work , assessing their peers-home or classroom work and if their students evaluate themselves differently, they were asked to specify the technique(Q41 **and** Q42);whether they have ever realized that the tests they administer to their students should take another formo,if "Yes" ,which ones(Q43 **and** Q44),and what form should they take instead(Q45).

Section 7: Teachers' Suggestions (Q46)

In this section the writer wanted to collect the teachers' suggestions and recommendations in the hope to develop an adequate battery of tests for the first year students of English, Faculty of Languages and Human Sciences, University of Béjaia

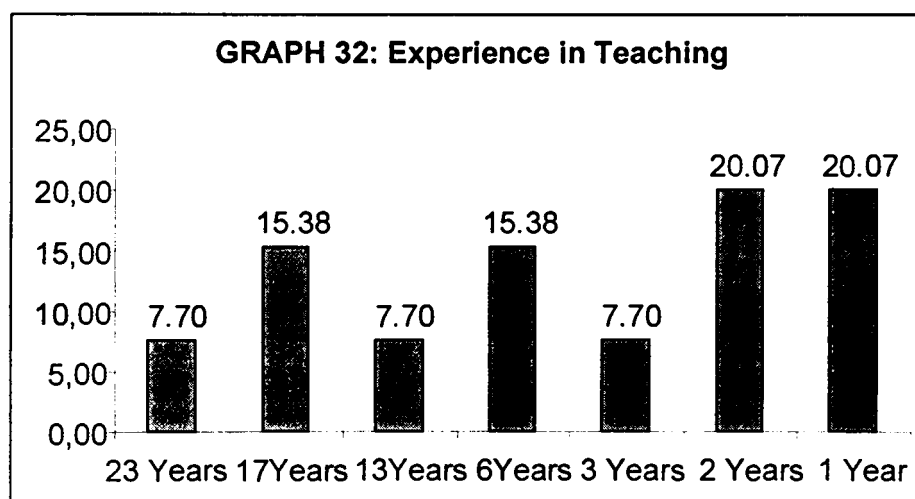
3- Analysis of the questionnaire.

Section 1: General Questions Question

1: -How long have you been teaching?

Table 37: Experience in teaching

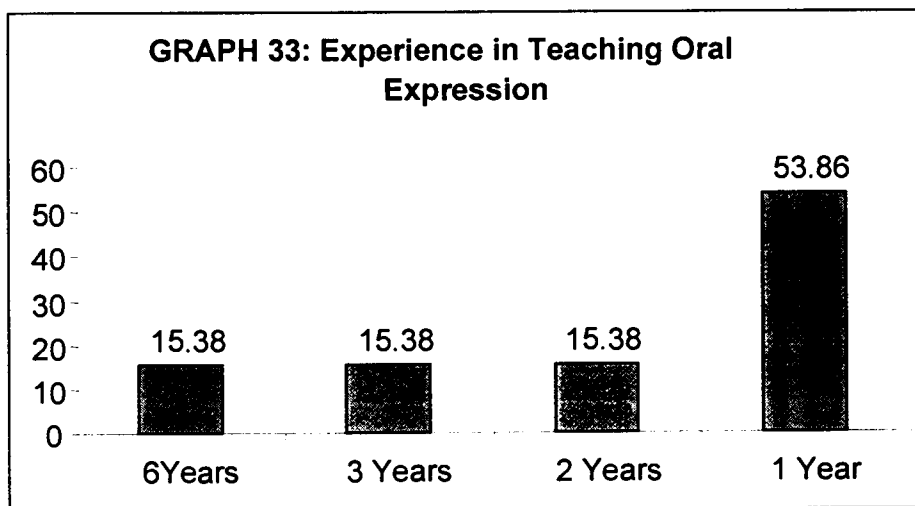
Years	Number	%
23 Years	01	07,70
17Years	02	15,38
13Years	01	7,70
06Years	02	15,38
03 Years	01	7,70
02 Years	03	23,07
01 Year	03	23,07
	13	100



Through the results obtained, we can notice that 53, 84 % of the questioned teachers have no more than 3 years experience in the field of teaching. Two of them i.e. 15, 38 % have 17 Years experience, and the most experienced has been teaching for 23 years.

Question 2: How long have you been teaching the oral expression?

Years	Number	%
06 Years	02	15,38
03 Years	02	15,38
02 Years	02	15,38
01 Year	07	53,86
Total	13	100

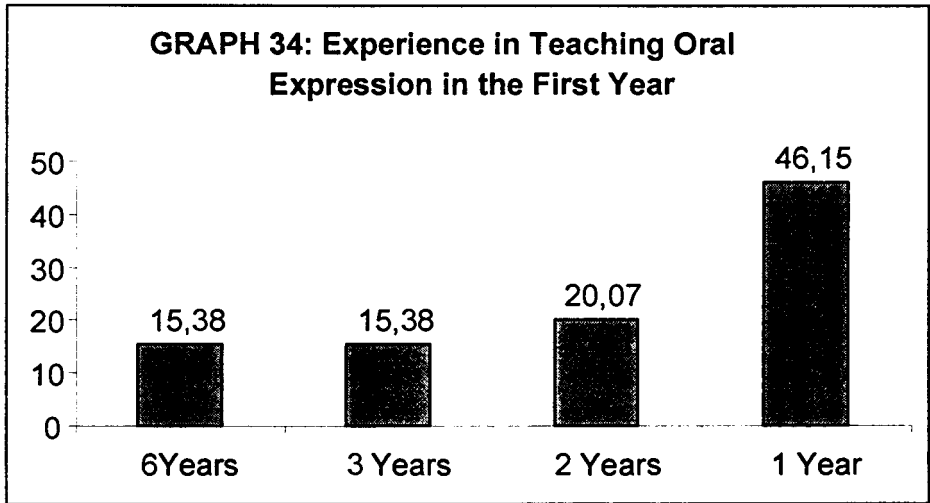


We can notice that the teaching staff of the Department of English enjoys a very short experience in developing the students' oral skill. The majority of teachers (seven out of thirteen) have only one year experience. This may let us think that they have just graduated from the university, and this lack of experience may have an impact on the quality of teaching.

Question 3: How long have you been teaching oral expression in the first year?

Table 39: Experience in Teaching Oral Expression in the First Year

Years	Number	%
06 Years	02	15,38
03 Years	02	15,38
02 Years	03	23,08
01 Year	06	46,16
Total	13	100



We can find out that nearly half i.e. 46, 16 % of the teachers of the first year have no more than one year experience in developing oral expression, and the most experienced who are two among the staff have an experience of six years. This can be explained by the instability of part time teachers who generally come from different "Lycées".

Question 4: Do you have any special training to teach oral expression?

Through the data collected from this question, we realize that a very small number (02) of the teachers of oral expression (15.38 %) have a special training to enhance the oral proficiency. This training was done in different private language schools of the city. This is not surprising because we know that the majority of the teaching staff is recruited as part time. Such results can be considered as motivating reasons to develop the idea of autonomy of learning and self-assessment in our learners.

Section 2: Principles of Testing

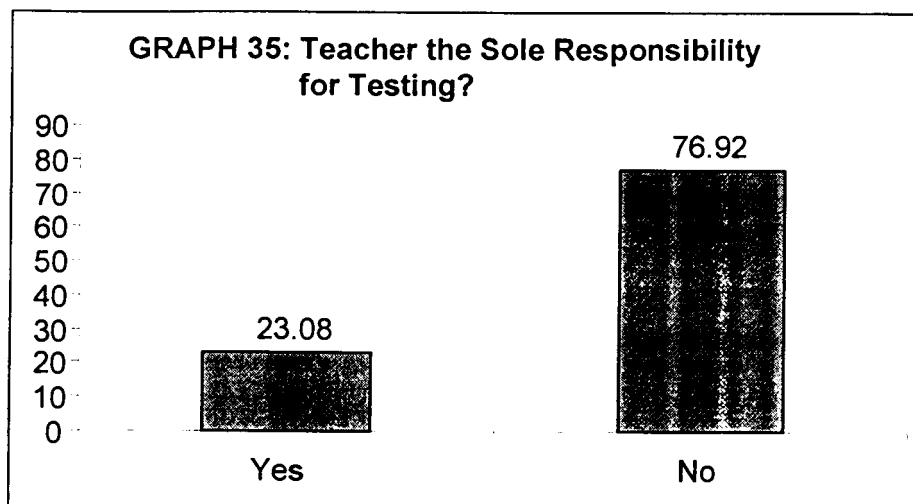
Question 5: Is the teacher the sole responsible for testing?

YES

NO

Table 40: Teacher the Sole Responsible for Testing?

Options	Number	%
Yes	03	23.08
No	10	76.92
Total	13	100



The results show that 76.92% of the informants (ten teachers out of thirteen) said that the teacher is not the only responsible for testing. We believe that such an opinion is really positive in the sense that it will involve and encourage learners to share responsibilities in the process of learning.

Question 6: Please, explain why.

The 10 teachers who answered "No", to this question argued that: students must also be involved in the assessment process and that the teachers must act and react as facilitators and advisors. As for the 3 other teachers who answered "Yes", they did not provide any explanation.

Question 7: Testing is:

a-Checking student's performance

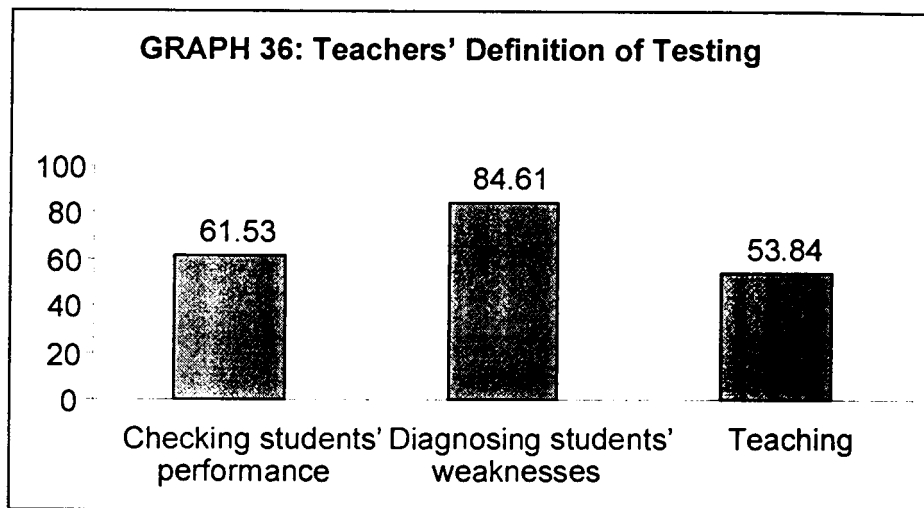
b-Diagnosing students' weaknesses.

c-Teaching

e-Other: Please, specify

Table 41: Teachers' definition of Testing

Options	Number	%
Checking student's performance	5	38.46
Diagnosing students' weaknesses	6	46.16
Teaching	2	15.38
Total	13	100



One can find out that the respondents have opted for one option at a time while defining testing, 46.16% opted for "Diagnosing students' weaknesses". In option "e", where we have asked the teachers to specify any other understanding of the term "Testing", (four teachers) specified that testing is also assessing the teachers' methodology of transmitting knowledge.

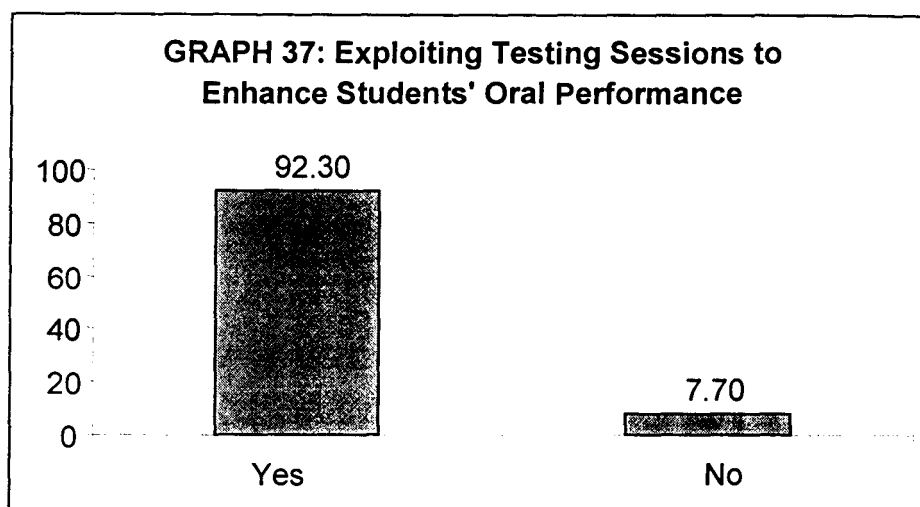
Question 8: Do you fully exploit your testing sessions to enhance your students' oral Performance?

Yes

No

Table 42: Exploiting Testing Sessions to Enhance Students' Oral Performance.

Options	Number	%
Yes	12	92.30
No	01	7.70
Total	13	100



Except one questioned teacher (among the thirteen) answered negatively. We believe that such an attitude is positive, because our students need to be encouraged to converse and interact in the target language.

Question 9: if "Yes", please, explain how?

Nine of the teachers who answered positively the question 8, explained that they manage to enhance their students' oral proficiency by creating a friendly environment, varying the topics, giving them opportunities to choose the subjects and things they like to talk about and share. As for the three others, they did not give any argument for their "Yes".

Question 10: If "No", please, say why?

The only teacher who answered "No" did not provide any justification for his answer.

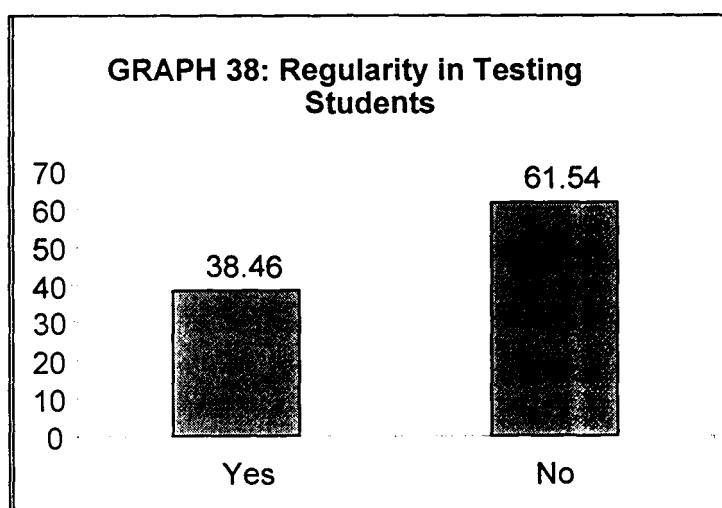
Question 11: Do you regularly test your students?

Yes

No

Table 43: Regularity in Testing Students

Options	Number	%
Yes	05	38.46
No	08	61.54
Total	13	100



Only five teachers (out of thirteen) said that they regularly test their students. Eight responded "No", three of them, justified, though they were not asked to, evoking time constraints.

Question 12: if "Yes", please, say why?

For those who answered "Yes", (5 of them) they explained that they test regularly their students because it is important to create in the learners readiness and possibility to be measured at any time during the process of learning. On the other hand, two of them said that regularity in testing allows both teachers and learners to know their strengths and weaknesses and try to do better.

Section 3: The Testing Procedure

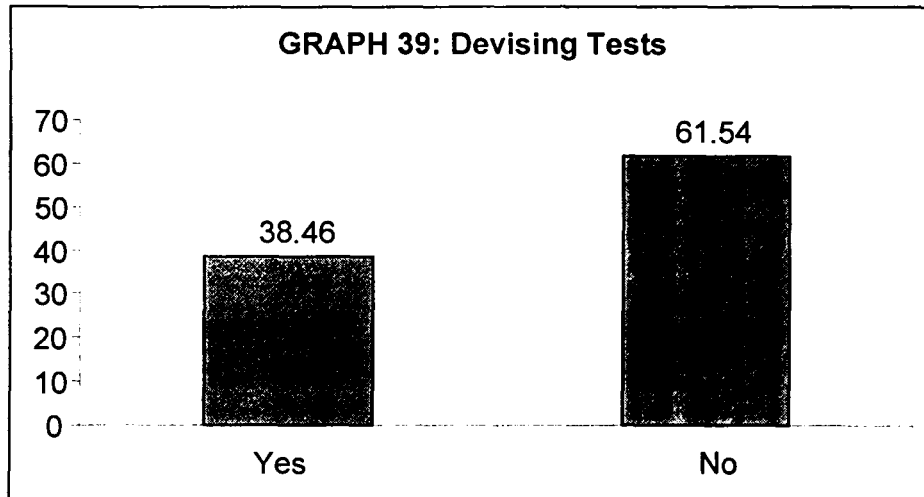
Question 13: -Do you write your own tests?

YES

NO

Table 44: Devising Tests

Options	Number	00
Yes	05	38.46
No	08	61.54
Total	13	100



Five teachers only write their own tests which represents 38, 46 % of the whole population under study. This may be due to the lack of experience of the teachers of the Department and the complexity of the nature of oral tests.

Question: 14: If "No", where do you choose them from?

a - A test book

b- A colleague provides them

c-Other: Please, specify

Among the eight teachers who said that they do not write their own tests, six of them specified that they choose them from a test book; two did not give any precision. We can note that no one of the informants said he/she gets the test from a colleague. This may imply to a certain extent the lack of coordination and cooperation within the teachers of the department.

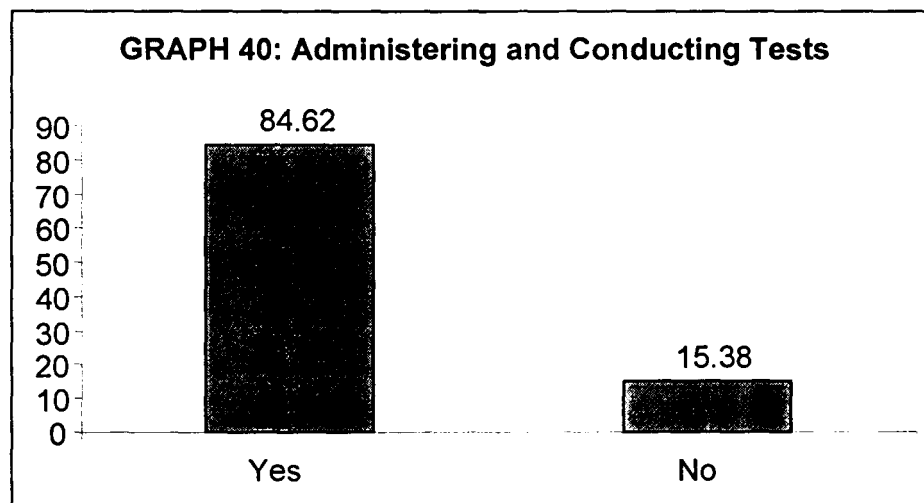
Question: 15 Do you administer and conduct your oral examinations alone?

Yes

No

Table 45: Administering and Conducting Tests

Options	Number	
Yes	11	84.62
No	02	15.38
Total	13	100



The results have shown that only two teachers said that they do not administer and conduct the oral examinations alone. This implies that they involve some other colleagues in the process. We consider involving other colleague teachers in the testing process will reduce to a certain extent subjective judgement and improve the reliability of the test results.

Question 16: Please, say why.

Among the 11 teachers who answered "Yes", two of them did not justify their answer. Four said that they have no choice but they do not agree with the existing system. Two said that: students feel at ease with their teachers when being tested. Two said that they administer the examinations alone because they know the students better .One of them said that each teacher is the only one to know the level and the abilities of his students. Through the data collected from this question, one may think that the arguments provided by teachers who do the tests themselves are not scientific but rather subjective, because it is accepted that the more raters we have in an examining board, the more reliable and objective the test will be. The two other teachers who answered "No" said that they invite other colleagues to take part in the examining board.

Section 4: Basic Considerations in Testing

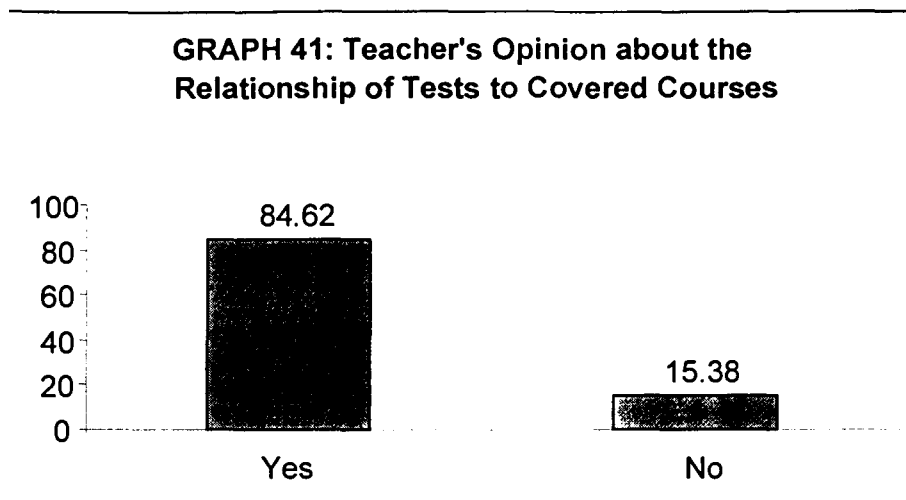
Question: 17- Are the tests you conduct related to the topics you covered in the 'niirce?

Yes

No

Table 46: Teachers' Opinion about the Relationship of Conducted Tests to Covered Courses

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



Eleven teachers (84.62 %) out of the total of thirteen confirmed that the tests they conduct are related to the topics they covered in the course, this may imply that the learners are given more chance to express themselves in the target language about familiar topics.

Question 18: If "Yes", is it:

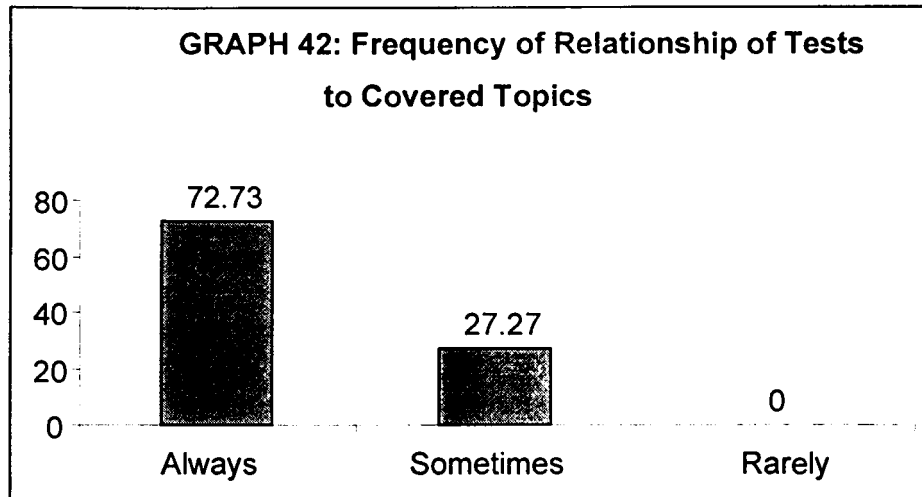
a- Always

a- Sometimes

a- Rarely

Table 47: Frequency of Relationship of Tests to Covered Topics

Frequency	Number	
Always	08	72,73
Sometimes	03	27,27
Rarely		
Total	11	100



Among the 11 among eleven teachers who answered "Yes", eight of them said that the tests they conduct are always related to the topics they have already covered. We think this will offer good opportunities to the learners to communicate in English.

Question 19: If "No", please, say why.

The two teachers who answered "No" to question 17, one **of them did not** justify his answer and the other one said that his/her choice is justified **by the** fact that he/she wants to encourage the students' reflexion and creativity.

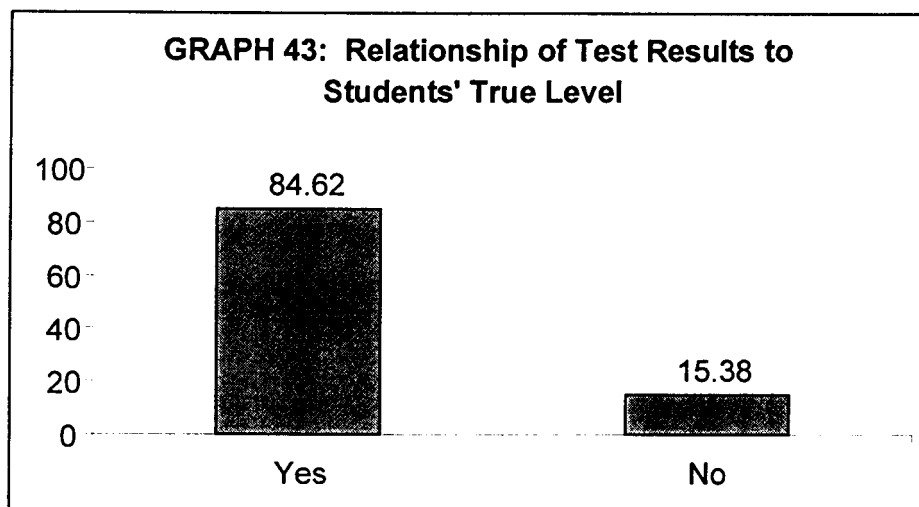
Question 20: Do the results obtained from your tests reflect your students' true level?

Yes

No

Table 48: Relationship of Test Results to Students' True Level

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



Eleven informants (84.62%) said that the results obtained from their tests do reflect their students' true level. This may ensure and encourage learners to work harder and try for better results.

Question 21: If "yes", is it:

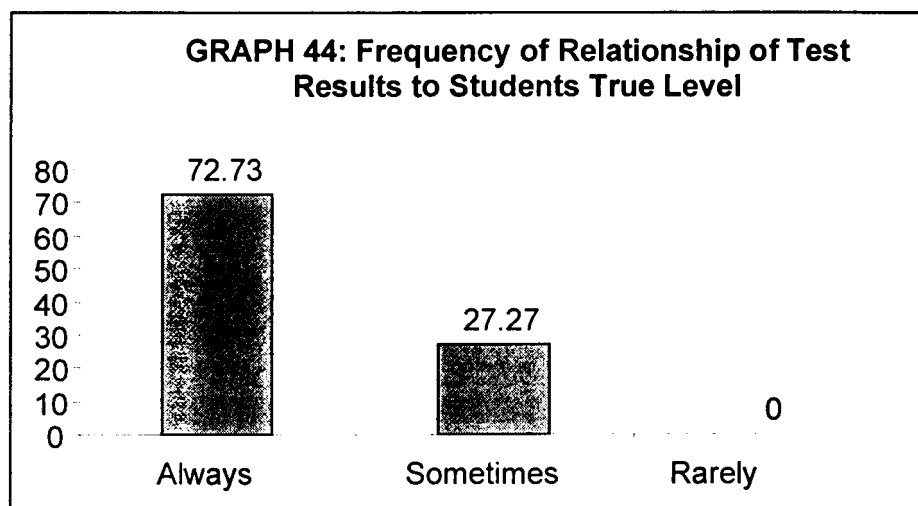
a- Always

b- Sometimes

c- Rarely

Tahla AQ • Fi rannanrov Ralatinchirs nf Tact Racnitc tn Qtnrlanc Trna t .aval

Frequency	Number	%
Always	08	72,73
Sometimes	03	27,27
Rarely		
Total	11	100



Amongst those who answered "Yes" eight mentioned that the obtained results always reflect their students' true level. This can be considered as positive results because this will reduce students' anxiety and frustration.

Question 22: If "No" _ do you give:

a- Another test

b- Remedial work

c- Other: Please, specify:

The two teachers who answered that the results obtained do not reflect their students' true level argued that they do not only give a remedial work but try to adopt another testing method.

Question 23: Are the test activities set at an appropriate level of difficulty?

Yes

No

All the informants confirmed that the test activities are set at an appropriate level of difficulty.

Question 24: If "Yes", is it:

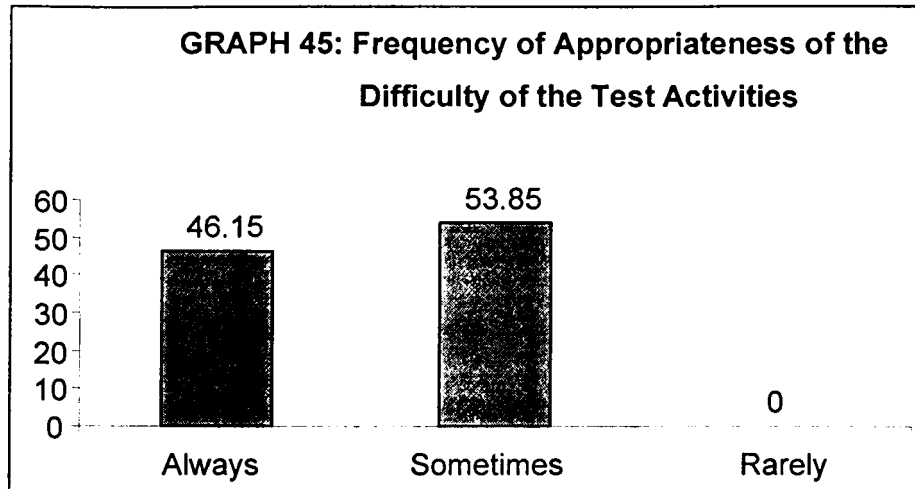
a-Always

b -Sometimes

c-Rarely

Table 50: Frequency of Appropriateness of the Difficulty of the test Activities

Frequency	Number	%
Always	06	46.15
Sometimes	07	53.85
Rarely		
Total	13	100



Among the thirteen teachers who answered positively the question, six of them said that the level of difficulty is always appropriate and seven mentioned that it is sometimes. This may let us suggest that some teachers in charge of the population of students under study need to reconsider their testing method.

Question 25: Does the test discriminate adequately between the performances of candidates at different levels of attainment?

Yes

No

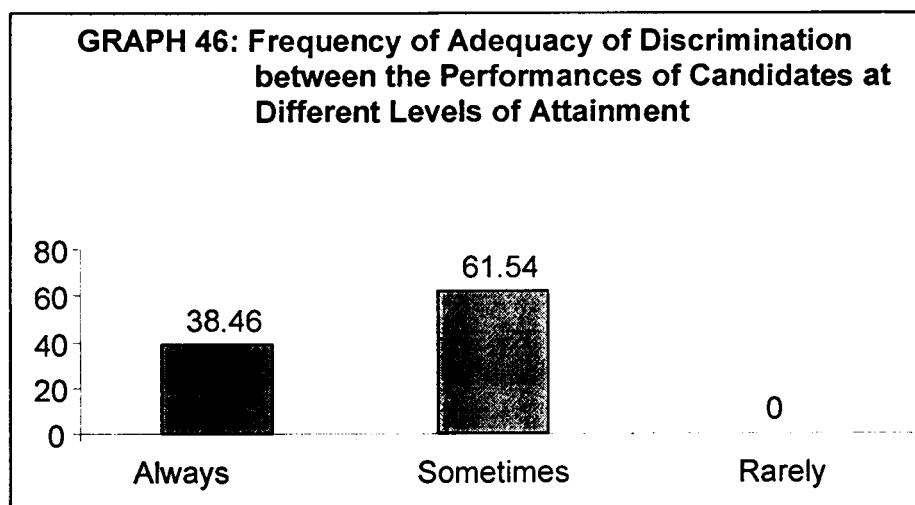
All the questioned teachers confirmed that the tests they deliver discriminate adequately between the performances of their candidates at different levels of attainment. This will give a chance to all students to practise appropriately the foreign language.

Question 26: If "Yes", is it:

- a- Always
- b- Sometimes
- c- Rarely

Table 51: Frequency of Adequacy of Discrimination Between the Performances of Candidates at Different Levels of Attainment

Frequency	Number	%
Always	05	38.46
Sometimes	08	61.54
Rarely		
Total	13	100



Five teachers (38.46%) said that the tests always discriminate adequately between the performances of candidates at different levels of attainment. Eight of them (61.54%) recognized that it is sometimes that tests discriminate between the performances of candidates. This may be due to the short of experience that the teachers have in devising tests for measuring the oral proficiency of their learners.

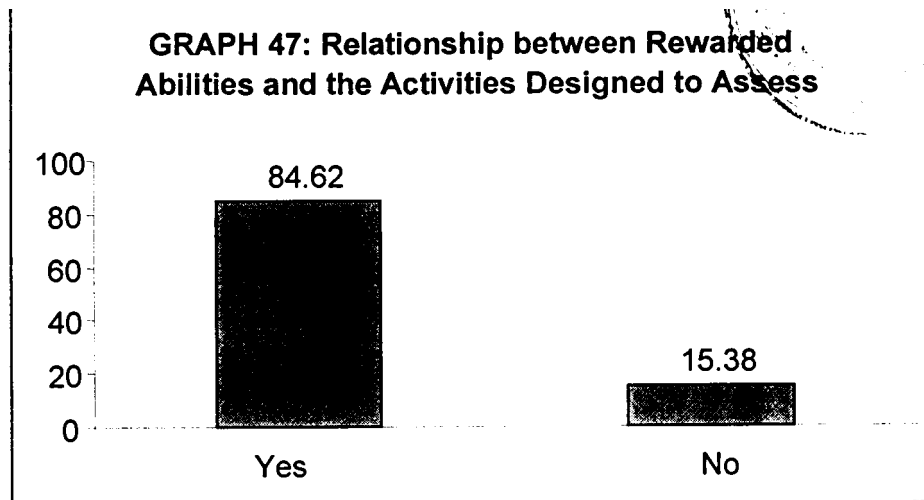
Question 27: Are the abilities being rewarded those which the activities are designed to assess?

Yes

No

Table 52: Relationship between Rewarded Abilities and the Activities Designed to Assess

Options	Number	ayo
Yes	11	84.62
No	02	15.38
Total	13	100 s



84.62% of the teachers answered positively, confirming that the abilities being rewarded are those which the activities are designed to assess. This is a good indication, in the sense that the learners are not trapped and misled while being asked to perform in English.

Section 5: The Marking Scheme

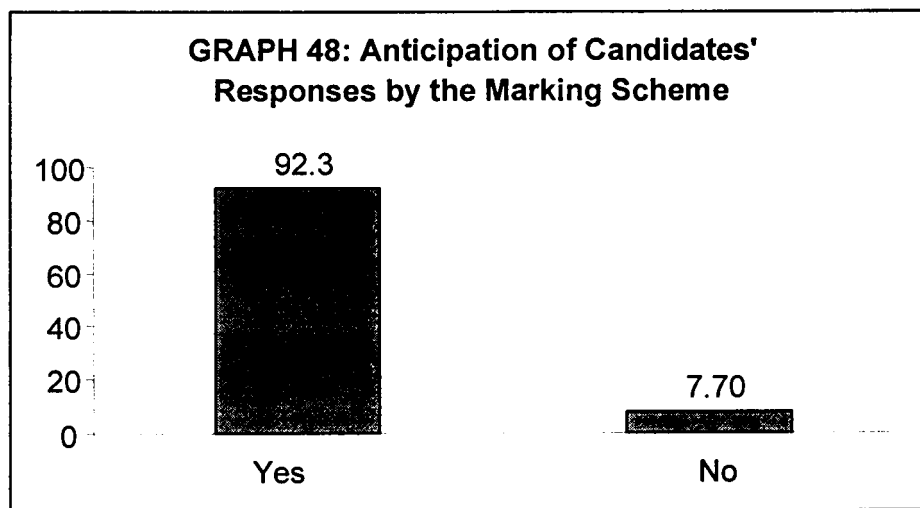
Question 28: Does the marking scheme anticipate responses of a kind that candidates are likely to make?

Yes

No

Table 53: Anticipation of Candidates' Responses by the Marking Scheme

Options	Number	%
Yes	12	92.30
No	01	07.70
Total	13	100



We can notice that all teachers except one answered that the marking scheme anticipates responses of a kind that candidates are likely to make. This will reduce the element of subjectivity and award the learners true grades.

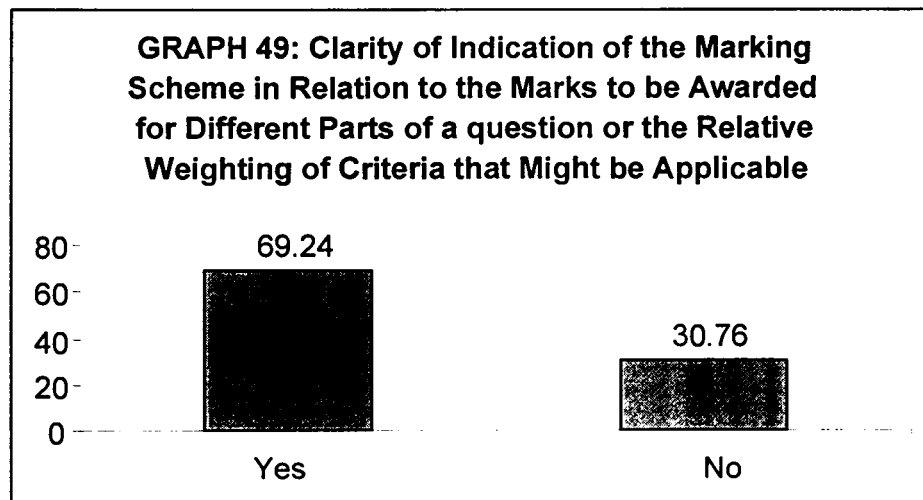
Question 29: Does the marking scheme indicate clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable?

Yes

No

Table 54: Clarity of Indication of the Marking Scheme in Relation to the Marks to be Awarded for Different Parts of Question or the Relative Weighting of Criteria that Might be Applicable.

Options	Number	%
Yes	09	69.24
No	04	30.76
Total	13	100



Nine teachers indicated that the marking scheme indicates clearly the marks to be awarded for different parts of a question and the relative weighting of criteria that might be applicable. We think that theoretically such results will improve the quality of students' test results.

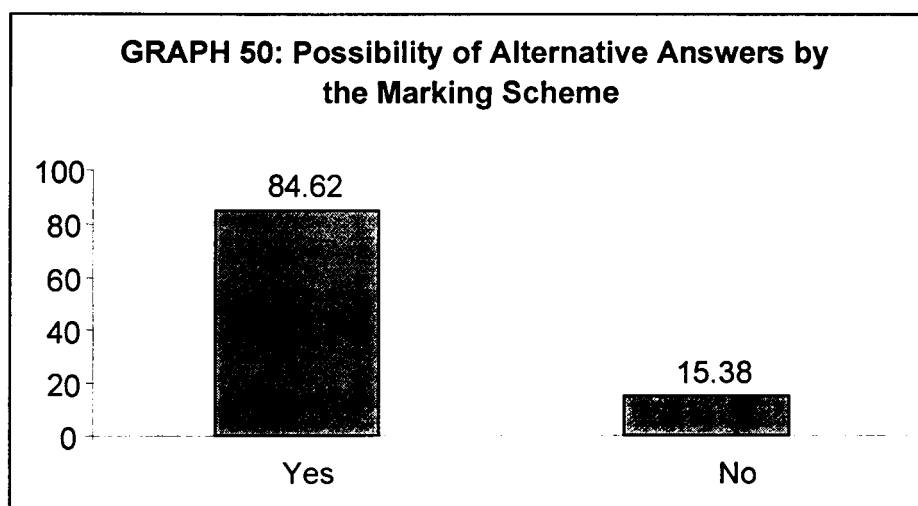
Question 30: Does the marking scheme allow for possible alternative answers?

Yes

No

Table 55: Possibility of Alternative Answers by the Marking Scheme

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



All teachers except two said that the marking scheme allows for possible alternative answers. This is positive attitude because this will allow flexibility and accepting the principle of achieving positive results differently.

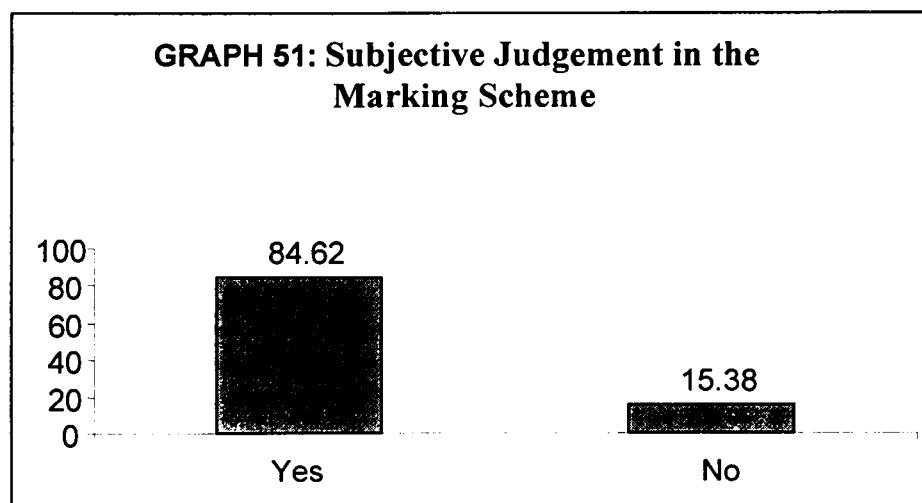
Question 31: Does the marking scheme, by specifying performance criteria, reduce as far as possible the element of subjective judgement that the examiner has to exercise in evaluating candidates' answers?

Yes

No

Table 56: Subjective Judgement in the Marking Scheme

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



A large majority of teachers (84. 62%) said that the marking scheme, by specifying performance criteria, reduces as far as possible the element of subjective judgement that the examiner has to exercise in evaluating candidates' answers. This will allow us to obtain results that reflect the true level of the students.

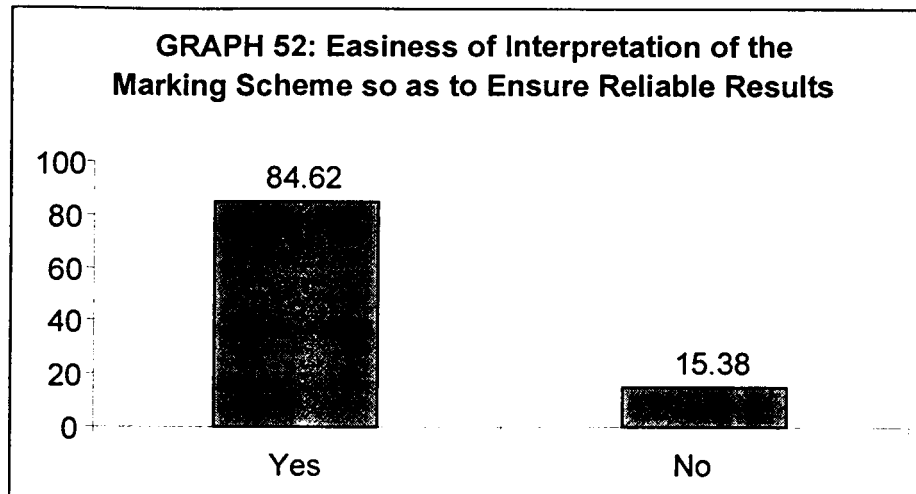
Question 32: Can the marking scheme be easily interpreted by a number of different examiners in a way which will ensure that all mark to the same standard?

Yes

No

Table 57: Easiness of Interpretation of the Marking Scheme so as to Ensure Reliable Results.

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



We can notice that the results obtained from question 31 and 32 are similar in terms of percentage (**84. 62%**) for "Yes". Such information is interesting in the sense that it obeys the norms of testing.

Question 33: Which language aspects do you test? Specify the order of importance (1 for the most important one to 5 the least important one.)

a- Meaning

b- Grammar

c- Pronunciation

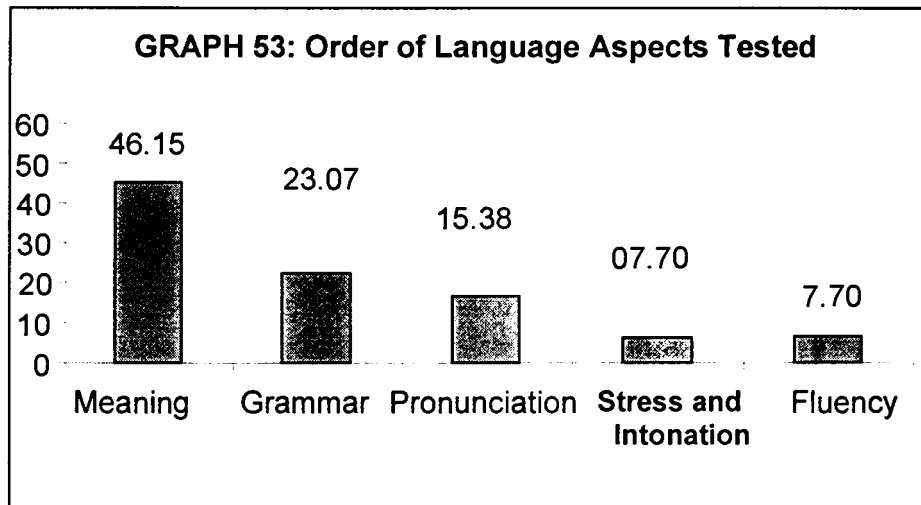
d- Stress and intonation

e- Fluency

f- Other: Please, specify:

Table 58: Order of Language Aspects Tested

	Language Aspects Tested	Number	00
1	Meaning	06	46.15
2	Grammar	03	23.07
3	Pronunciation	02	15.38
4	Stress and intonation	01	07.70
5	Fluency	01	07.70
Total		13	100



One can notice that 46.16 % of the teachers concentrate on the meaning aspect while testing their learners. Our opinion is that all aspects must be taken into consideration when testing the oral proficiency, but specific attention to any aspect of the language will depend on the objective of the test.

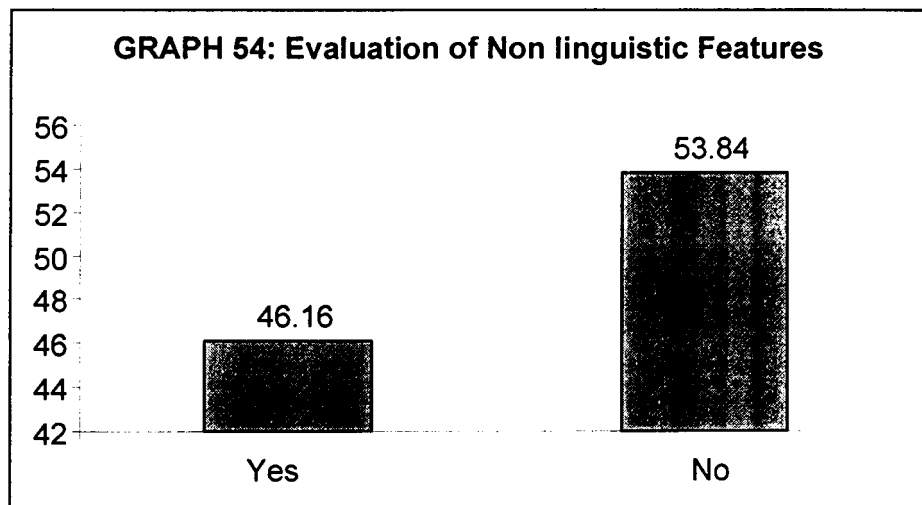
Question 34: Do you evaluate non linguistic features, such as attitude, gestures, way of dressing?

Yes

No

Table 59: Evaluation of Non linguistic Features

Options	Number	%
Yes	06	46.16
No	07	53.84
Total	13	100



We have found out that 53.84% of the questioned teachers do not evaluate non linguistic features, such as attitude, gestures, and way of dressing. We think that raters should take into consideration at least to a certain extent these paralinguistic features while measuring the oral proficiency.

Question: 35 If "Yes", is it:

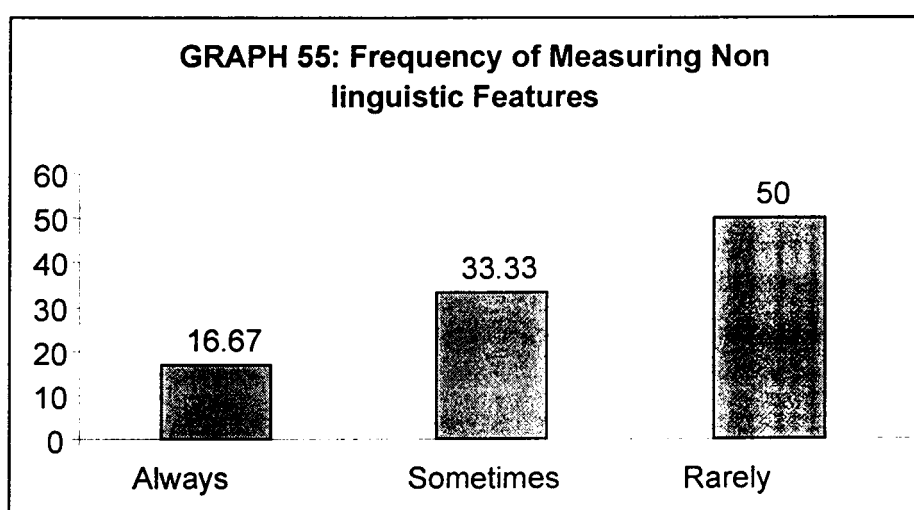
a- Always

b -Sometimes

c-Rarely

Table 60: Frequency of Measuring non linguistic Features

Frequency	Number	%
Always	01	16,67
Sometimes	02	33,33
Rarely	03	50,00
Total	06	100



We can find out that among the six teachers who answered "yes" to question 34, three of them (50%) said that it is rarely **that they take into** consideration the non linguistic features in their evaluation. We believe that one can not deny totally the non linguistic features of learners, because they can be a good indicator to diagnose the students' abilities.

Question 36: If "No", please, explain why.

Three teachers out of the seven said that only speaking is supposed to be examined. Three others said that they do not believe that such features are important. One said that students are old and wise enough and free to dress.

Section 6: Oral Tests

Question 37: Which testing techniques do you use to evaluate your students' oral proficiency?

a-Multiple Choice Questions

b-Oral Interview

c- Group work

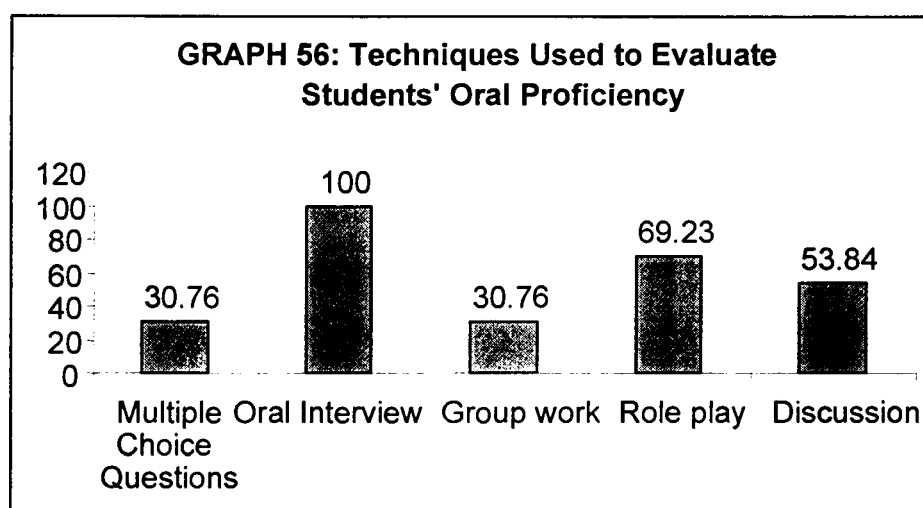
d-Role play

e-Discussion

f-Other: please, specify:

Table 61: Techniques Used to Evaluate Students' oral Proficiency.

Used technique	Number	%
Multiple Choice Questions	03	23.07
Oral Interview	08	61.53
Group work		
Role play	01	07.70
Discussion	01	07.70
Total	13	100



The results from question 37 indicate that 61.53% of the teachers use the technique of oral interview in testing and developing their students' oral proficiency. The choice of this technique by the majority of the teachers may be explained by the fact that conducting an interview does not require much preparation and investment. We note that each teacher has opted for one single testing technique.

Question 38: Have you heard of self –assessment?

Yes

No

All teachers of the oral expression have said that they have already heard of Self- Assessment. We find the answer interesting because it goes with the writers' objective to introduce this new vision of assessment.

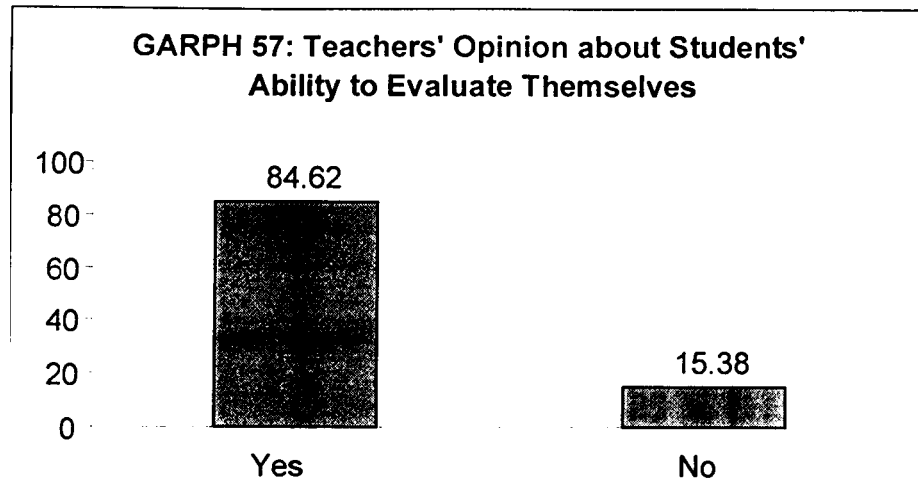
Question 39: If "Yes", do you think your students are able to evaluate themselves?

Yes

No

Table 62: Teachers' Opinion about Students' Ability to Evaluate Themselves

Options	Number	%
Yes	11	84.62
No	02	15.38
Total	13	100



Eleven teachers (84, 61%) believe that their students can assess themselves. Only two showed their pessimism for such assessment. We believe that once they experience this new alternative of assessment, they will no doubt discover another positive way of doing things differently.

Question 40: Please, explain why.

The explanations provided by the teachers who think that their students can not assess themselves are that their students need preparation, an argument that we do share because our learners really need special training for such assessment. They have also said that their students do not have the required capacities so as to be involved in such responsibilities. As for those who answered "Yes", ten of them said that this type of assessment will develop in the learners the sense of criticism, self-reliance and autonomy; one of them did not provide any answer.

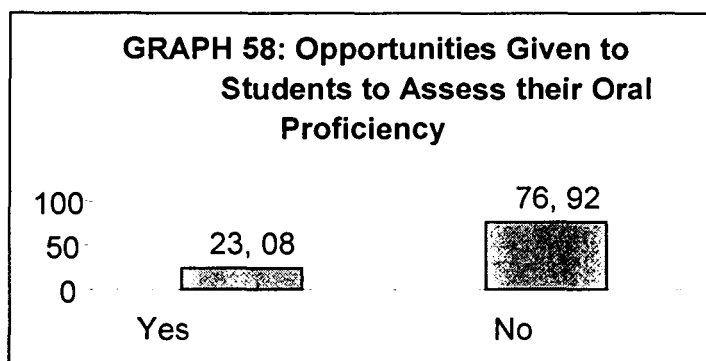
Question 41: Do you give opportunities to your students to evaluate their oral proficiency?

Yes

No

Table 63: Opportunities Given to Students to Assess their Oral Proficiency

Options	Number	%
Yes	03	23.08
No	10	76.92
Total	13	100



Only a small percentage (23, 07%) of teachers give opportunities to their students to evaluate their oral proficiency. This is probably due to the fact that the teachers themselves have not been trained and prepared for such new responsibilities. Here, we notice a contradiction with the answer to Q 39, where 84.62% of the teachers said that their students can self-assess but we realize that the learners are not given this opportunity.

Question: 42 If "Yes", is it through:

a- Assessing their own home or class work

b-Assessing their peers-home or classroom work

c-Other: Please, specify.

Two teachers said that students usually assess their peers-home or classroom work. One said that it is through assessing their peers-home or classroom work, provided with a model of the corrected work.

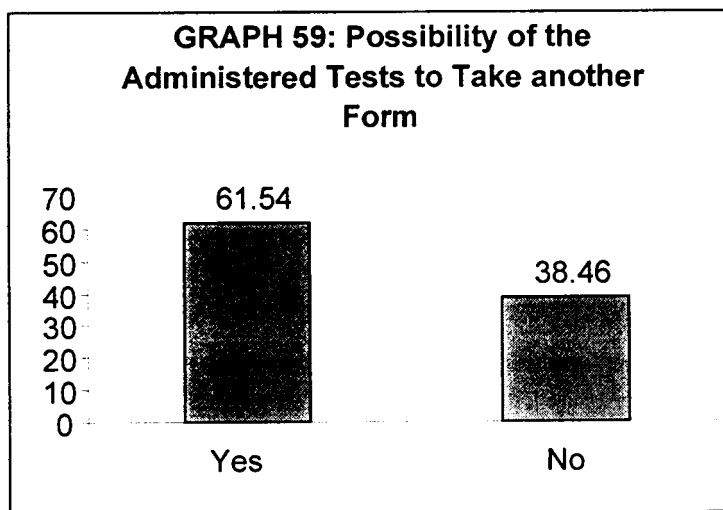
Question: 43 Have you ever realized that the tests you administer to your students should take another form?

Yes

No

Table 64: Possibility of the Administered Tests to Take another Form

Options	Number	%
Yes	08	61.54
No	05	38.46
Total	13	100



From the collected answers we have found that 62% of the questioned teachers believe that the tests they conduct should take another form. We think that this opinion consolidates the writers' objective to introduce and propose "the Portfolio" as an alternative way of assessment.

Question: 44 If "Yes", which ones?

We have noticed that all the respondents to this question said that Group work, Filling the Gaps and Multiple Choice Questions should take another form.

Question: 45-What form should they take instead?

The eight teachers who said that they realize the tests they administer should take another form have proposed students' self-assessment and independent guided research projects.

Section 7: Teachers' Suggestions

Question 46: Would you please add any suggestion or recommendation?

The suggestions and recommendations of the teachers of the oral expression are:

All of them admitted that they often encounter some difficulties, not only in developing the Oral Proficiency of their learners, but in adopting the adequate testing techniques to measure the ability to communicate in the target language. They justified their difficulties by the fact that they lack good experience and efficient training. However, they have expressed a strong wish if they can have a specialized intensive training so as to overcome all the faced difficulties.

Nine teachers said that their pedagogy and methodology of teaching and testing will be improved considerably if there is a better coordination and cooperation between the concerned teachers.

Conclusion

The needs of first year LMD students do not vary much, because the great majority of the informants, (more than **50 %** see table: 8 and 9) expect their oral testing sessions to become situations that will free them from nervousness and hesitation. This new expected situation will enable them to perform and communicate in English orally. They seem to be aware of the importance of learning English. For this reason, we highlight the urgent need for the implementation of the new LMD programme which will foster in our learners the ability to use English appropriately and naturally during their period of study and eventually during their job careers.

The teachers' questionnaire has shown that there is a strong need to train the newly inexperienced recruited teachers for different credits they will be in charge. We have also realized that there is an urgent need to set and consolidate habits of working in collaboration with the totality of the teachers of the oral expression so as to meet real needs of the students.

CHAPTER SEVEN

PORTFOLIO EVALUATION

Introduction	258
I- Developing a Portfolio	258
1-Direrctions to the Students	258
2- Students' Portfolios	262
II- Scoring Portfolios	288
1- Criteria of Portfolio Assessment	288
2- The Rating Scale	294
III-Evaluation of the Experiment	297
Conclusion	299

CHAPTER SEVEN

Portfolio Evaluation

Introduction

After the analysis of the students' and teachers' questionnaires, we have realized that foreign language teachers in the Department of English use many different types of assessments in their classrooms. In addition to multiple-choice tests, teachers use, oral interviews, role-play, discussion, group-work, but no portfolio assessment. As a researcher, we have taken into consideration the students' and the teachers' opinion as for the production, implementation and assessment of students' portfolios. Since the portfolio is evidence to learners of their own efforts, progress, and achievements, we believe that it is a suitable tool for self-assessment. To evaluate students' progress and achievement using these authentic assessments, teachers need an accurate rating system. Scoring rubrics, which combine assessment criteria with a rating scale, can help make an assessment reliable and valid

I-Developing a Portfolio

1- Directions to the Students

The oral proficiency of the population under study which is a sample of 29 L.M.D first year students (a group out of the 12 groups which makes a total of 372 first year L.M.D students) has been evaluated through "a Portfolio". The choice of this particular group was done on the only basis that the researcher is their tutor. Right from the beginning, we went through different steps introducing and familiarizing the learners with the idea of "a Portfolio", explaining that it is a collected, organized, annotated body of work, produced

over time by the learner, which demonstrates progress towards specific objectives. Their Portfolio work will represent their development as speakers and language users over time. We also explained that the time they spend developing their portfolios is not time lost to instruction and learning. It is through portfolio development that students use the skills and knowledge that are part of the curriculum. In this study we opted for a selective "Portfolio" which is a selection of their work that best represent their progress in this course as users of English.

Right at the beginning of the experiment and before conducting the test, we explained very clearly to the students involved in the study that what they need to do is to include artifacts in their portfolios that evidence their development with respect to the study of English and their reflections on that development. For example:

- the notes you have prepared for class discussion as you were listening to an evening news,
- a description of a group work you have been involved in class;
- formal discussion or piece of writing;
- your reflections on the meaning of a particular chapter in a reading;
- a recording of you speaking English, reciting a poem;
- a newspaper article you have read about a topic related to English and your reflections on it;
- a grammar exercise you have completed;
- the results of a (tutoring) session or a portfolio conference;
- an email exchange with a classmate in which you have discussed aspects related to class or English;
- an example of notes you have taken in class in an English for Specific Purposes (ESP) lecture;

- comments on any of the activities you have been involved in for the Course Project;
- comments on corrections you have received from your tutor.
- various versions of the Course Project;
- a comparison between the first and final corrections you have made in your portfolio or E-mails;
- presentation of a biography of an important figure.

The learners will realize that the contents of the portfolio are almost endless!

We specified to the students what, and how much, has to be included in the portfolio. We offered possibilities to the students to include topics that enable self-expression and independence. We did not miss to tell them that they should not necessarily place their best work in their portfolios, but the work that most clearly represents their development. We also explained that it is not acceptable to simply put copies of assignments into their portfolios, and if they did they will not receive a favourable evaluation. Rather, they were encouraged to reflect on their work and comment on why they think particular artifacts represent their development and progress. This means that for each sample of their work that they place in their portfolio, they must also place an accompanying piece in the portfolio that explains in detail how the first piece represents their development.

We have also clarified to the students that comparison and reflexion are crucial elements of a portfolio, and the need for clear and attractive presentation is really recommended and appreciated. For example, they may choose to place the first and second drafts of their first recording of a passage or a discussion in their portfolio. Alongside these two artifacts they must place their reflections on their development from the first version to the final version. They might include aspect like giving concrete examples of how they began to introduce themselves

during a summer course in English speaking country, using different expressions for introducing themselves, telling about an important scientific, historical, artistic, political figure. They might also discuss content realizations that they had in the course. We have proposed to the students, examples of the types of aspects they might place in a portfolio, but the portfolio is not limited to these artifacts and the principle of a portfolio is that they can place anything of interest in there.

An important aspect of portfolio assessment is student/teacher conferences (tutoring sessions). These sessions have been an integral part of portfolio assessment right at the beginning of the semester. Graves (2000) explicates the purpose of a conference by stating that during this time, the teacher elicits information from the student rather than issuing directives about errors on work included in the portfolio. Goal setting by the students is an important part of the conferencing/tutoring process. Research and experience have shown that students often set standards in conferences for evaluating themselves as learners and for setting goals that are different from their teachers'. Many researchers, among them Graves 2000 and Smith, 1989: suggest that conferences encourage and teach students to reflect critically on their work. Reports from these same researchers reveal the importance of conversation in the conferencing process. They also suggest that both teachers and students are learners during the process. Additionally, our experience has showed that student teacher conferences based on reflection and self-assessment and the portfolio preparation, provide students the opportunity to actively participate in their own learning and give teachers an insight into the student as a learner.

2- Students' Portfolios

As already mentioned in this chapter, there were 29 first year LMD students who were involved in the project of preparing an individual portfolio. We admit that it was not an easy task for learners to produce evidence and self assess, simply because reflection and self assessment do not come naturally to people who have had no practice in it.

It is true that many students at first felt uncomfortable with the portfolio conception and self-evaluation, thinking that it is the job of the teacher to evaluate; some teachers (mostly those who are still in favour of the traditional way of evaluation and others who do not want to have more pedagogical responsibilities) were also reluctant at the beginning to this new vision of learning and testing, evoking time constraints. But, as we were personally determined and convinced that this alternative is a safe way for our students who are willing to grow as active learners and set a path for future learning, we did not give up. Instead, we provided them with the necessary Guidance, Support and encouragement through tutoring sessions.

It is worth mentioning that the data compiled on individual students has provided a clear picture of each student's development through the various work samples and products collected. Through the produced and presented portfolios, we have noticed the possibility to determine growth, areas of weakness, and areas of strength. We have also been informed about the student's background, interests, and goals through his / her portfolio, conversations, observations and the tutoring sessions. In short, we became acquainted with the entire group involved in this study. Thus, contrary to traditional testing, which typically provides only a set of numbers, alternative assessment documents a story for every student – what is the ultimate goal of evaluation if not to give us the knowledge to be able to reflect on, discuss, and assist a student's journey

through the learning process? Alternative assessment gives us the power to do all three.

So, this is how we managed to help the group of students under study to demonstrate that they can be positive learners, by producing and self assessing a motivating personal, work. We have selected three Portfolios produced by the students with the guidance and under the supervision of their teachers. Through these productions, we wanted to show how we can develop the learners' autonomy of learning and self confidence. These sample Portfolios are presented here; they reflect examples of a good Portfolio, an average portfolio and the portfolio which is characterized by some difficulties.

SAMPLE 1 :

***The differences between
British and
AMERICAN ENGLISH***

Outline

1- Introduction:

- a- The origins of British English
- b- The origins of American English
- c- The influence of naive languages

2- The increase British English and American English have had on each other
3- The differences :

The spelling differences

- a- Short and long A
- b- Stress differences
- c- Possible confusions
- d- Lexical differences
- e- Grammatical differences
- f- The Gotten /Got differences

4- Conclusion

1- Introduction :

a- The origins of British English:

Before the Anglo-Saxon invasion, the languages spoken by the natives belonged to the Celtic family, introduced by people who came to the British Isles around 1000 B.C. The Roman arrived in 43 B.C. and left the British Isles in 410 B.C. By the year 1425, English became widely used in England, in writing and in speech.

At the end of the 18th century, English was not yet identical to the one used today. Many words were spelled the same as the ones used now, but had different meanings. Pronunciation and word stress were also held different from the way they are in present day English.

b- The origins of American English:

American English has more native speakers than any other form of English. It is a flexible ever-changing language. From the French and the Spanish, who at one time also owned parts of North America, American English inherited a vast vocabulary. Even today, new immigrants bring with them new languages and new customs, and foreign words enter the language all the time.

c- The influence of the native languages:

When the first English speakers arrived in America, they met native Americans who spoke many different languages and dialects. Although some of these languages have died out, some words found their way into modern

American English. This is because of the early European trappers and traders learned words from the natives, so they could do business with them. But even the new languages influenced the native's one. The immigrants adopted the language spoken in America in addition to their own; as a result, they developed a new language, and created new words.

2- The increase that BrE/AmE have had on each other:

There is a considerable variation in the case of American and British English, but here are no accurate estimates for the number of points of contrasts. Recent decades have seen a major increase the two models have had on each other, especially American English on British.

The influence of US films and television has led to a considerable passive understanding of much American English vocabulary in Britain, and some of this has turned into active use, especially among younger people. The reverse pattern is less obvious, but British films and TV programmes are seen sufficiently often in the USA to mean that a growth in awareness of UK vocabulary should not be discounted.

The regional dialects surveys of both countries, several of which have only recently begun to publish their findings, are bringing to light huge amounts of lexical distinctiveness.

3- The differences :

a- The spelling differences :

In the American spelling, advertising has been responsible for some unusual spellings. The idea of spelling words is not a new one. In the 1780s, Noah Webster, the most famous of all American dictionary-makers supported

the idea of eliminating the letters that were not pronounced in a word. He has also been in favour of changing the order of letters in words so that they matched the pronunciation. Thus, his American speller is chiefly responsible for the spelling differences in American and British English.

These spelling differences were noted as one of the main sources of variation in the world press. Advertising has been responsible for some unusual spellings. Several of them are productive, applying to large numbers of words such as:

<u>British English</u>	<u>American English</u>
Colour	Color
Catalogue	Catalog
Aeroplane	<i><u>Airplane</u></i>
Cheque	Check
Inflexion	Inflection
Manoeuvre	Maneuvre

B- Short and long A:

In England, one accent has traditionally stood out above all others in its ability to convey association of respectable social standing and a good educational. This 'prestige' accent is known as Received Pronunciation, or RP which has many words in 'long a' that are pronounced with 'open a' in GE which means General English. These are some examples:

Advance	example	overdraft
Aunt	fasten	pass
Basket	France	pastor
Castle	grass	plant

Disaster

laugh

sample

c- Stress differences:

There are many words whose stress varies between the two accents. Some of them can be grouped into patterns, such as those ending in —ary / -ory (ex: secretI, laboratQ),or —et (ex: ballet, beret) , which attract stress on the final syllable. Some of the words vary, depending on their sentence position, as in " princess Ann is a princess" .

This is one of the areas where American influence on British English has been particularly strong, and probably lot of words can be heard in the UK (United Kingdom) these days with the American stress pattern, especially spoken by younger people.

Received Pronunciation

General American

Address

Address

Advertisement

Advertisement

Ballet

Ballet

Cafe

Cafe

Cigarette

Cigarette

Controversy

Controversy

d- Possible confusions:

The transcriptions use the same symbols in different ways, partly because of different views about the best way to analyse the vowel system, and partly because of the differences between British and American English.

• /ɑ/ in the British systems does not appear as a separate phoneme. In F&R (the system used by Victoria Fromkin and Robert Rodman) it is used in such words as *dog*, reflecting more directly the way this vowel is articulated further forward in the mouth. This is a major point of possible confusion for British — trained students casually reading an American transcription, for they risk interpreting /lag/ as log instead of log.

In addition, the same /ɑ/ symbol is used by F&R in such words as *father*, *calm* and *car*, again reflecting the typical sounds of these vowels in American English, whereas the British systems use /ɑ:/ - an important difference between the two sound systems.

• /e/ in F&R refers to the vowel in such words as *sue*, whereas the British systems show diphthongal nature of this sound as /ei/ or /eI/ .

Thus, /met/ refers to *mate* in F&R, but to *met* in Gimson in an introduction to the pronunciation of English (A.O Gimson is a British phonetician).

• /ʌ/ in the British systems refers only to the vowel in such words as *seen*. In F&R it is also used for the vowel in such words as *bird*.

• /o/ in F&R refers to the vowel in such words as *so* British students used to a diphthongal transcription would there for be likely to interpret /kot/ as *cot* rather than *coat*.

• /t/ in BrE as in *water* and *often* is pronounced *wa?er* and *of?en* in AmE, since /r/ in AmE as in *confront* is pronounced in BrE *conf?ont*

e- Lexical differences:

In the lexical differences, the words used in BrE and AmE are different lexically, but have the same meaning. These are some examples:

American English

British English

Allowance	Pocket money
Apartment building	Block of flats
Baggage	Luggage
Biscuit	Scone
Bookstore	Bookshop
Draft	Conscription
Elevator	Lift
Fall	Autumn
Gasoline	Petrol
Vacation	Holiday
Lawyer	Solicitor
Mail	Post
Purse	Hand bag

f- Grammatical differences:

There are relatively few grammatical differences between educated BrE and AmE .A leading reference grammar notes regional trends affecting only C250 points in morphology or syntax, with many of those affecting individual items (ex: irregular verbs), and very few being general points of syntactic construction. (AmE examples precede BrE in the following paragraphs, for grammatical terminology.

- In the verb phrase, AmE refers "have" to "have got" for possession (do you have the time? Vs have you got the time?), answers also tend to vary (I don't Vs I haven't).AmE prefers such forms as burned to burnt, and there are some special past tense, forms AmE sometimes uses a simple past tense, where BrE has a present perfect (I just ate Vs I've just eaten), will won't is generally found for shall / shan't.

- There are several differences in propositions and adverbs, such as AmE I'll go momentarily (Vs in a moment), real good (Vs really good)....

g- The Gotten/Got distinction:

Gotten is probably the most distinctive of all the AmE / BrE grammatical differences but British people who try to use it often get it wrong. It is not simply an alternative for have got. Gotten is used in such contexts as "they've *gotten* a new boat (=obtain)", "they've *gotten* interested (=become)" and "he's *gotten* off the chair (=moved)", but it is not use in the sense of possession (have).

AmE does not allow "I've *gotten*", but "I've *got*" as in informal BrE.

The availability of gotten however mean that AmE can make such distinctions as the following:

"They've got to leave" which means they must leave, since "they've gotten to leave" means they've managed to leave.

4- Conclusion :

Despite the fact that, today, there are many differences between British and American English, both of them are fascinating languages. It's really important to learn English which is a world language and that becomes nowadays the chief means of communication between nations.

- There are several differences in propositions and adverbs, such as AmE **I'll** go momentarily (Vs in a moment), real good (Vs really good)....

g- The Gotten/Got distinction:

Gotten is probably the most distinctive of all the AmE / BrE grammatical differences but British people who try to use it often get it wrong. It is not simply an alternative for have got. Gotten is used in such contexts as "they've ***gotten*** a new boat (=obtain)", "they've ***gotten*** interested (=become)" and "he's ***gotten*** off the chair (=moved)", but it is not use in the sense of possession (have)

AmE does not allow "I've *gotten*", but "I've *got*" as in informal BrE.

The availability of gotten however mean that AmE can make such distinctions as the following:

"They've got to leave" which means they must leave, since "they've gotten to leave" means they've managed to leave.

4- Conclusion :

Despite the fact that, today, there are many differences between British and American English, both of them are fascinating languages. It's really important to learn English which is a world language and that becomes nowadays the chief means of communication between nations.

SAMPLE 2:

*Biography of Diana
Princess of Wales*



Content

The Biography of Princess Diana:

- 1) Childhood and Teenage years
- 2) Marriage and Family
- 3) Public Role
- 4) Death



Childhood and Teenage Years:

BIRTH NAME: **Diana Frances Spencer** was born July 1, 1961 (Sun in Cancer, Moon in Aquarius), at Sandringham, Norfolk, England. Diana was born at Park House, the home that her parents rented on the estate of **Queen Elizabeth II** and where her childhood playmates were the Queen's younger sons: **Andrew** and **Edward**. She was the youngest of the three daughters of **Edward John Spencer** and **Frances Ruth Burke Roche**, who were later divorced (1969).

HER EDUCATION: From 1966 until 1974; Riddles worth Hall Prep School, Norfolk; 1974-77; West Heath near Sevenoaks, Kent; 1977-78; finishing school at the Institute Alpin Videmanette in Rougemont, Switzerland, where she perfected the French language and became a competent skier. In 1979, upon returning to London, she lived with three female friends in a rented apartment at Coleherne Court in South Kensington. For a while she looked after the child of an American couple. She then worked as a kindergarten teacher at the Young England School in Pimlico, London.

Marriage and Family:



Her friendship with **Charles**, the **Prince** of Wales, first in line to the British throne, grew in 1980, and on the 24th day of February 1981 it was announced that Lady Diana Spencer was to marry the Prince of Wales. She was the first Englishwoman to marry an heir to the throne in over 300 years. They were married on July 29, 1981, at St Paul's Cathedral, London. Watched by 750 million people worldwide on television. She had two sons; **Prince William**, born on June 21, 1982 and **Prince Henry** (Harry) on September 15, 1984, both at St Mary's Hospital, Paddington, and London.

Eleven years after her marriage and after many reports and accusations concerning marital infidelity and emotional alienation between the two, the couple were officially separated on December 9, 1992, and remained a great deal detached from each other ever since. In November 95, Diana admitted having a love affair with her riding coach James Hewitt (who wrote a book about the affair). On February 28th, 1996, the Princess announced her readiness to officially divorce Charles. In November 1995, the Princess gives a television interview during which she spoke of her unhappiness in her personal life. The prince and princess were divorced on 28 August 1998.

Public Role:



After her marriage, the princess of Wales quickly became involved in the official duties of royal family. Her first tour with the Prince of Wales was a three days visit to Wales in October 1981. In 1983 she accompanied the prince on a tour of Australia and New Zealand, and they took the infant prince William With them. Prince William with Prince Harry Again joined the Prince and princess of Wales at the end of their tour to Italy in 1985. Other official overseas Visits undertaken With Prince include Australia, Brazil, India, Canada, Nigeria, Cameroon, Indonesia, Spain, Italy, France, Portugal and Japan.

2.2-A Continuing Debate

In spite of this accumulated evidence, broadcasters and scientists continue to debate the link between the viewing TV violence and children's aggressive behaviour. Some broadcasters believe that there is not enough evidence to prove that TV violence is harmful. But scientists who have studied this issue say that there is a link between TV violence and aggression, and in 1992, the American psychological association's task force on television and society published a report that confirms this view. The report, entitled *big world, small screen: the role of television in American Society* shows that the harmful effects of TV violence do exist.

3-What parents can do

While most scientists are convinced that children can learn aggressive behaviour from television, they also point out that parents have tremendous power to moderate that influence. Because there is a great deal of violence in both children's programming, just limiting the number of hours children watch television will probably reduce the amount of aggression they see.

In addition:

Parents should watch at least one episode of the programs their children watch. That way they'll know what their children are watching and be able to talk about it with them. When they see a violent incident, parents can discuss with their child what caused the character to act in a violent way. They should also point out that this kind of behaviour is not characteristic, not the way adults usually solve their problems. They can ask their children to talk about other

ways the character could have reacted, or other non-violent solutions to the character's problem.

Parents can outright ban any programs that they find too offensive. They can also restrict their children's viewing to shows that they feel are more beneficial, such as documentaries, educational shows and so on.

Parents can limit the amount of time children spend watching television, and encourage children to spend their time on sports, hobbies, or with friends; parents and kids can even draw up a list of other enjoyable activities to do instead of watching TV.

Parents can encourage their children to watch programs that demonstrate helping, caring and cooperation. Studies show that these types of programs can influence children to become more kind and considerate.

As we have already mentioned earlier, *the three sample portfolios* we have included in this research are students' own productions accomplished with the guidance and under the supervision of their tutors.

The first Portfolio deals with the Differences between British and American English, including its origins and the different influences English has undergone. The Oral presentation was considered of a good standard and the topic judged very interesting because it is related to some of the credits taught to first year LMD students (Linguistics and phonetics).

The second Portfolio is related to the Biography of the Lady Diana, Princess of Wales, including her education, marriage, her public role and her tragic death. The board of examiners and the researcher considered the content of the portfolio very interesting (illustrating a moment of British history) and the oral presentation was evaluated "average"

Sample portfolio three is an illustration of the effects of seeing violence on television. The data collected by the students emphasises particularly, the aggressive attitude of children and teenagers who, watch such programs. The board of examiners admitted that the topic is interesting, but the learner has shown some serious difficulties in the Oral presentation. (Very frequent hesitations in speech, frequent and serious grammatical and word order mistakes, and poor vocabulary and word choice.

II-Scoring the Portfolios

1- Criteria of Portfolio Assessment

At the end of the experiment, there was a need to evaluate the portfolios in their entirety. We have noticed that evaluating and grading an organized collection of student work, self-assessments, and reflections can be a complex process, and at a certain time we felt that this is one of the most challenging and difficult tasks in the portfolio process, because the Subjectivity of judgements on our part as raters is often cited as a concern in this type of assessment (Bateson, 1994).

To reduce subjectivity and succeed in the task of evaluating portfolios, we realized the need to prepare an organized system of assessment and monitor students' portfolios. Evaluating the portfolio as a whole takes careful planning. One of these steps is setting a clear purpose. We determined who will evaluate **the** portfolios and how they will be evaluated. We also specified how the portfolio will be assessed. We have even acquainted the students with the scoring guides/rating scales that will be used before.

In addition to setting the purpose, we adopted a rubric to bring structure and reliability to the evaluation. This rubric combines the criteria on which the portfolios will be assessed with a rating scale. The criteria are clear and measurable statements of what the learner should be able to do or what he/she should show in the portfolio. For instance, to document progress in communicating in the target language, the criteria might be:

- communicates information about a variety of topics through clear, organized, and well-developed oral presentations;
- Uses appropriate and varied vocabulary;
- Uses learned grammar structures accurately;

- Uses a process communicating approach.

In order to use a portfolio as self-assessment for evaluation purposes, we believe that questions to guide the self-assessment must be provided. Sample questions could include the following:

- To what extent did you achieve your goals in learning since you have been involved in the production of your portfolio?
- To what extent did you improve your speaking and methodology? List some of the problems you faced while speaking and how you solved those problems.
- To what extent has your knowledge of vocabulary improved? List the new words you have learnt from your portfolio.
- To what extent has your confidence in using English improved?

By referring to their portfolios in answering these questions, learners have concrete evidence of their performance and are not forced to rely on their intuition and possible bias about their performance or ability.

In order to help the students to self —assess their productions, we have proposed some assessment scales that can be used in the context of listening, speaking and some rubrics for measuring learners' portfolio presentation. We believe that the suggested scales will be useful for the teachers in charge of LMD students and others who are willing to be involved in this type of assessment.

Table 65: Student Self-Assessment of listening and Speaking

Student

Date.....

Group.....

How would you rate yourself as a listener and speaker of English? Rate yourself from 1 (a beginner) to 5 (an expert).

My Rating	Beginner		Intermediate		Expert
	1	2	3	4	5
COMPREHENSION (understanding when people speak)					
• Outside of class					
• In class				•	
FLUENCY (Speaking smoothly without hesitating)					
• Outside of class					
• In class					
VOCABULARY (understanding and using specific words)					
• Outside of class					
• I n c l a s s					
PRONUNCIATION (saying the words clearly)					
• Outside of class					
• In class					
GRAMMAR (using the rules of the language)					
• Outside of class					
■ In class					

Table 66: Proficiency Guidelines: A Self-Assessment of Speaking

Student.....

Date.....

Group •

Think about how you use language; decide how well you can do what is asked in the target language. Tick the appropriate column

When SPEAKING, I can :	Not so well	Ok	Quite well	Really well
Ask questions				
Answer questions				
Introduce myself				
Make a request				
Talk a bout myself				
Talk a bout my family				
Apologize when I do something wrong				
Complain when something is not right				
Tell a story				
Describe something				
Tell the facts about an event				
Discuss something of personal or public interest				
Support someone's opinions				
Explain in detail				
Express what could happen				
Participate in formal conversations				
Discuss school subjects ,LMD system				

Table 67: Rubric for measuring students' Oral proficiency

Student.....

Date.....

Group.....

SPEAKING

Level	Functions	Forms
Novice	<ul style="list-style-type: none"> ■ Expresses basic courtesies • Handles elementary needs • Asks simple questions • Makes statements 	<ul style="list-style-type: none"> • Uses some basic vocabulary such as objects, places, and family terms • Produces isolated words, phrases, or short sentences • Shows some signs of spontaneity • Has frequent errors
Intermediate	<ul style="list-style-type: none"> • Handles limited interactive, task oriented, and social situations • Talks simply about self and family members • Participates in short conversations about personal history and leisure-time activities • Asks and answers questions 	<ul style="list-style-type: none"> • Uses vocabulary to express the most elementary needs • Combines and recombines elements with some connected discourse • Hesitates and pauses, causing strained fluency • Pronounces often with first language influence • Communicates meaning in conversations
Advanced	<ul style="list-style-type: none"> • Narrates and describes with connected discourse • Satisfies requirements of everyday situations and school routines • Elaborates, complains, apologizes • Discusses topics of personal and current interest • Uses communicative strategies such as paraphrasing 	<ul style="list-style-type: none"> ■ Generally shows fluency and ease of speech • Links sentences together smoothly • Uses vocabulary to communicate finer shades of meaning
Superior	<ul style="list-style-type: none"> • Participates in informal and formal conversations on practical, social, academic, and abstract topics • Support opinions and hypothesizes using native-like strategies 	<ul style="list-style-type: none"> • Produces some sporadic errors but no patterns of errors, are evident • Uses technical, low-frequency vocabulary

Table 68: Rubric for Assessing Portfolio Oral Presentation

APPROPRIATENESS OF LANGUAGE	FLUENCY OF SPEECH	GRAMMAR AND WORD ORDER	VOCABULARY AND WORD CHOICE
5-extremely appropriate 4-appropriate 3-adequate 2- inadequate 1-inappropriate	5-like a native speaker 4-very smooth 3-some hesitations 2-frequent hesitations 1-choppy	5-no mistakes 4-occasional but unimportant 3-occasional;more serious 2-frequent and serious 1-meaning obscured	5-appropriate for situation and level 4-very good 3-adequate 2-poor 1-insufficient for situation, etc

2-The Rating Scale

The rating scale measures the extent to which the learner has met the criteria by using descriptors such as rarely/sometimes/often/always or yes/no. We have chosen the rating scale based on the type of criteria and the method of grading. For example, when we wanted to qualify students' progress in writing their portfolios, we have decided to choose a rating scale with descriptors such as no improvement/some improvement/significant improvement. Before the assessment period begins, we have developed and discussed the scoring rubric and have shared it with the students. We have modified the rubrics according to changes in instruction during the experience. At the end of the experience, we have used the final version of the rubric to evaluate the portfolios.

The rubric provides structure for the evaluation and strengthens the accuracy of the assessment; however, during the assessment of the portfolios we tried to monitor the contents of the portfolios to ensure reliable scoring. For instance, since we were evaluating portfolios for evidence of progress, we tried to make it sure that students' samples are the same genre and same degree of difficulty in order to compare format, vocabulary, and grammar. The samples were written under similar conditions.

To prepare for the final evaluation, we felt the need to decide how to grade the portfolios. This task has presented a specific challenge to us as teachers first, because it is recognized that testing Oral proficiency has an intrinsically subjective nature; achieving reliability in speaking tests is to some extent difficult (Alderson, Clapham & Wall, 1995; Brown & Hudson, 2002; Carroll & Hall, 1985; Hughes, 1989; Weir, 1990). The problematic nature of reliability in oral assessment may stem from inconsistencies among raters; yet, we were convinced that using more than one rater will definitely contribute to the reliability of oral ability testing. To ensure the reliability of oral assessment,

we agreed that two or more raters as suggested by (Bachman & Palmer, 1996; Berkoff, 1985; Brown & Hudson, 2002; Genesee & Upshur, 1996; Henning, 1987; Hughes, 1989; Madsen & Jones, 1981; Norris, Brown, Hudson, & Yoshioka, 1998; Underhill, 1987; Weir, 1990) assess the same student performances and combine the grades they assigned to the same student.

Second, the rating scales used for assessing oral performance have presented some difficulties related to the number and clarity of the categories in the scale. As Davies et al. (1999: 53) explain, a rating scale is a framework that serves as a "scale for the description of language proficiency consisting of a series of constructed levels against which a language learner's performance is judged". In order to minimize the possibility of different interpretations of scale descriptors by different raters (Alderson et al., 1995), language categories have been clearly defined. Moreover, the categories included in a rating scale and the different weightings awarded to different categories depend on categories which are regarded as relatively more important than the others according to a particular language program (Brown, 1996; Carroll & Hall, 1985; Hughes, 1989; Underhill, 1987). In assessing oral performance, establishing a clear assessment procedure and using explicit criteria were really essential to increase reliability (Hughes, 1989; Underhill, 1987; Weir, 1995).

Third, components of oral ability itself are not defined clearly (Hughes, 2002; Madsen, 1983), which leads to problems in choosing components to measure and using the test to provide feedback to students. When designing oral assessment tasks, we have realized that it is essential that each speaking test should have a clear purpose (Alderson et al., 1995; Bachman & Palmer, 1996; J. D. Brown, 1995; Carroll & Hall, 1985; Cohen, 1994, 2001; Graves, 2000; Hughes, 1989; Weir, 1995); therefore, the nature of the assessment criteria to be used depends on the purpose of a test. Among the general purposes of

assessment, such as assessing progress, proficiency, and achievement, three common assessment purposes in speaking tests are important in this study, namely, distinguishing among strong and weak students, giving instructors feedback on the effectiveness of their instruction, and giving students feedback on their learning process. Brown (1995) asserts that tests whose contents are the main language focus in language classes let raters receive feedback on effectiveness of their instruction and give their students feedback on their learning process. Thus, raters are likely to find rating scale categories whose objectives are taught in their classes more assessable and, therefore, more important than those whose objectives do not reflect course content. We have also noticed that the assessment of oral ability can be negatively affected by the discrepancy between test content and instruction. That is to say, instruction and tests should be in harmony with each other; assessment criteria should be incorporated into the syllabus and considered in lesson planning procedures (Hughes, 2002; O'Malley & Valdez Pierce, 1996).

How do we assign a single score to a collection of a student' work that focuses on growth and development? Our experience has shown that grading portfolios is beneficial because it encourages students to take them seriously and to complete the assignment. On the other hand, assigning a single score to the portfolios may reduce their effectiveness in capturing progress over time.

Deciding how to grade the portfolios depends on the teacher's purpose for using them. It also depends on the teacher's personal philosophy and on the learning environment. For example, in this research, through the presentation of portfolios, we assess quality of performance and progress in speaking. We have also graded the portfolios for organization and adherence to guidelines to ensure that students complete the assignment. Thus, the individual teacher with the collaboration of the colleagues has to determine the best approach to grading based on the teaching environment and on the objectives.

III-Evaluation of the Experiment

After the first couple weeks of experimenting on the portfolio, we realized that if we spent five minutes reflecting on the twenty nine portfolios each week, it added up to more than two hours per week for the class under study. We decided to stay with it and two aspects stuck in our mind. First, after two sessions of looking at the portfolios, which in the beginning included self-assessments, goal-setting, strategies information, and some diagnostic work, we felt as though we had gained more insight into these students than after a whole first semester of instruction without portfolios. We truly understood the concept of individualized instruction and saw how different students' needs were. We also felt that a strong bond of trust has been established between the students and ourselves. We were amazed at how willing students were to open up, that they did want someone to understand them and understand how they learned. Secondly, we realized that we could not pore over their portfolios every week. We looked at them in-depth on a monthly basis and put more responsibility on the students through self-and peer-reflections. It became easier over time-everyone got used to the routine. The portfolios also caused us to become purposeful and organized in our instruction, and this saved us time day-to-day. We cannot say that portfolios take less time than traditional paper and pencil scanned tests, but after seeing the benefits, we strongly believe that they can definitely serve as an alternative assessment.

An important part of portfolio evaluation is reflecting on the assessment process at the end of each semester. After evaluating the portfolios in their entirety, we have reviewed the evaluation procedure for validity and reliability. This was done by asking ourselves and finding out if the portfolios served the assessment goal, and if the evaluations were accurate and consistent. For example, if the teacher's and students' fixed objective was not met, then we need to determine why this is so. Perhaps, there were not sufficient samples of

student work to measure progress; or **maybe the rubric was not accurate enough, and the teacher became subjective in scoring the portfolio.** The teacher will want to **identify the weakness in the assessment and/or the instruction and remedy it for the next time.**

Finally, we can say that evaluating foreign language portfolios is a challenging task with many issues to consider. To succeed, we should set up an organized system of assessment by determining a purpose, creating a scoring rubric, and deciding how to grade the portfolios. Throughout the semester and with the good will of the teachers and the motivated students, we have updated the rubric as needed and monitored students' portfolios for appropriate work. As the end of the experiment approached, it was important to be sure to set aside enough time for the evaluation activity. After the final evaluations, we took time to review and revise our assessment plan. With preparation, the process has gone smoothly, and the assessment has contributed to students' learning.

We have suggested to the students to rate their current abilities using a rubric appropriate to the task performed (for example, descriptors, such as beginner, intermediate, expert or when I speak I can perform such and such activity with some ease or difficulty). Our students then have known the criteria with which they will assess themselves. For example, if the student is to evaluate how well he/she did in a unit on communication of biographical information, he/she needs to know that the goals to be met were being able to 1) give oral information on his/her own family members, 2) asking questions about someone else's family, 3) understanding a written description of a family. The student can then use a rubric (for example, I can do this, I can not do this) to assess his/ her ability to perform each of these tasks. Self-evaluation of a task can also be more detailed. For example, a student evaluating how well he/she is able to talk about her family can use criteria such as 1) speaking fluently with

few pauses, 2) using a wide variety of words personalized to his/ her family, and 3) using correct adjective/noun agreement.

Through our experimentation and while comparing the teachers' scoring of the portfolios and the students' self-assessment of their presentation, we reported a high correlation between self- evaluation and teachers' grading. Right at the beginning, we were somehow worried about the students who will be biased in their own favour while self assessing and consequently running the risk to distort the results. We rather noticed during the interview that some of the good students tended to under estimate the quality of their oral proficiency and, this is probably due to their awareness of what remains to be learnt .Whereas, we noticed a few cases among less good ones who ventured to overestimate their oral proficiency; this bias can be explained by the fact that these learners ignore their limitations and all the language violations they commit while performing.

Conclusion

This chapter is the result of the keen interest shown by the writer to introduce and implement in our educational system a new vision of learning and evaluating through a Portfolio. We understood that in order to optimize the results of our research, no element should be neglected. However, the necessary directions, suggestions and the scoring rubrics are provided to succeed in developing and scoring a portfolio. This new alternative of doing will free all pedagogical actors from the traditional and the daily routine, which hinders progress along the learning process.

CHAPTER EIGHT

RECOMMENDATIONS

Introduction	301
I- Designing a Portfolio	301
II- Answers to Teachers' Concerns about Portfolio Assessment	304
III- Answers to Students' Concerns about the Oral Tests	309
IV- Advice and Action Plan for Students	314
V- Advice to Teachers	316
Conclusion	318

CHAPTER EIGHT

RECOMMENDATIONS

Introduction

A significant amount of experimentation in new or refined methods of performance assessment is in progress. Much of the effort focuses on authenticity or realism of the assessments (tests), the standards against which to measure student performance, procedures for rating or scoring the new assessments, and training of educators in how to use and score them.

As already mentioned in Chapter 5, a portfolio might contain a number of performance tasks or assessments of those tasks. In many cases, performance tasks require construction, creation, description (written or oral), or other formats for task completion that lend themselves to portfolio inclusion. Often, performance tasks are quite structured in time and space.

I- Designing a Portfolio

In the elaboration of a portfolio, we suggest the following questions can form the skeleton of portfolio design. Designers may wish to add others that address uniqueness in their students or settings.

- What instructional goals, objectives, and outcomes do we want to measure?
- Which ones (goals/objectives/outcomes) are not now being assessed adequately by other means?

Concerning the answer to this question teachers need to know that we do not reinvent the wheel. If current assessment methods are adequate, why switch?

- How portfolio entries and their analysis are used to assess individual student progress over time and compare student accomplishment taking into account individual differences?

- What evidence of progress and/or accomplishment will be required? What evidence of progress/accomplishment will be allowed? Here, the question addresses the need for a consistent base of information from students and addresses issues of individual differences such as creativity, best effort, learning styles.

- Who will select entries? Why? We have to note that in some cases, teachers select all entries. In others, students build their portfolios within specific guidelines. Several researchers and developers recommend that both parties be contributors.

- What types of evidence can/will be accommodated in the portfolio? Why? This question was addressed in chapter five where it was stated that the type of portfolio, its storage and retrieval system, the area(s) of content involved in the assessment and the characteristics of the students have to be considered.

Underlying the questions above are a number of philosophical and measurement issues that need to be discussed and some agreement reached by the professionals in a program, before portfolio utilization is undertaken. Perhaps, the discussion is best facilitated by development of propositional statements such as those below which are offered for debate. They are a compilation of many of the premises found in the current literature on student portfolios.

- Portfolios can best be used to assess a student's ability to produce, perceive and reflect.
- Portfolio entries should be selected by both students and teachers by mutual agreement. Both parties have a stake in the teaching/learning process.
- In program assessment, portfolios provide insight into process as well as products and outcomes.
- Portfolios are best used to assess student development over time rather than to assess comparative accomplishments of students.
- Portfolios do little to accommodate learning styles unless students are encouraged to produce and submit diverse types of materials and products.
- Portfolio development and cooperative learning activities go hand-inhand, (The two can be easily related).
- If student reflection is desired, both self-critiques and teacher critiques of entries are required; so that teachers and students can compare them.
- Evaluation of portfolio contents requires at least two levels of organization: categorical organization of raw data/evidence and summaries or syntheses of available data.
- To score a portfolio, we must include at least two raters so as to reduce the risk of subjective judgement.

We have to remember that unfamiliar ways of teaching and assessment are potentially threatening and confusing to students. It is important to present the portfolio guidelines clearly, and to go over the guidelines periodically. Although all the guidelines - goals, content, timetable, etc. should be presented to the students orally so that they can discuss the procedure and ask questions, there should also be written guidelines to back-up the points discussed and for reference while preparing the portfolio. It is helpful to prepare these guidelines in a question-and-answer form

At the beginning, it is important to follow certain recommendations:

- We need to start small, by choosing one class and the objectives of the portfolio. We have to work out the portfolio process with the first class, and then add another class the next semester or the following year. Experience has shown that teachers are more likely to experience success if the portfolio system is manageable. Success the first time will give teachers the confidence to continue and expand in the future.

- We have to get students involved and made responsible for their learning. There is no need to do things for them that they can do themselves. Let students participate in selecting and organizing materials. Involvement leads to a sense of responsibility, a valuable trait for any student to have.

- We advocate the participation of teachers who are not involved in portfolio projects to take part. We should recruit and train native speakers of the language in the school to help. The native speakers gain professional interpersonal communication experience, and students increase exposure to authentic language.

- Students are required to work on their portfolios outside of class and asked to make comments on their work. However, in any case, we have to be sure to periodically review the portfolios in class for the sake of accountability.

II-Answers to Teachers' Concerns about Portfolio Assessment

We met with all the teachers of the department, and we came out with a number of questions to which we tried to provide some answers.

1. Q: I like the idea of portfolio assessment but isn't it a lot more work?

A: Indeed, it is hard work, but we are sure that spent efforts will be well rewarded by your students' increased motivation and involvement in their learning, and, in many cases, by improved achievement. The challenge of portfolio assessment lies not so much in the amount of work as in the organisation of the whole endeavour. Effective portfolio assessment requires planning in advance and keeping records; these will quickly become a habit and result in more efficient, professional work.

This work offers suggestions to help you plan and implement portfolio assessment systematically. In addition, teamwork will greatly reduce the burden on you. Do not over-extend yourself at the beginning: limit the scope of your first portfolio project to allow enough time for essential features of the process, such as tutoring and practice in reflection. Use portfolio assessment for a semester in one of your classes or require a limited number of entries before deciding how to go on. We should not forget, checking tests is laborious too, while the portfolios will save you some of that work.

2. Q: I am used to assessing my students totally on my own. With portfolio assessment, I see that the assessment process also includes students' evaluation of themselves and each other. How reliable are my students' evaluations?

A: It is true that students' self- and peer-assessments may not be very reliable at first - that is why training in this vital life-skill is very important. With time and practice, and especially given clear assessment criteria and individual guidance, students will learn to become better evaluators of their work. The main benefit of students' assessments will be in deepening their understanding of their own learning in order to improve their achievements. It is important to train students

in these skills and also to assess them, but we have to be careful not to make them into a significant component of the students' grades too soon.

3. Q: Is it fair to make everyone take part in portfolio assessment, especially when we are teaching a heterogeneous class?

A: We think that Portfolios are the ideal assessment tool for the heterogeneous class; for they are open-ended tools that help the teacher draw a **profile** of each student's strengths and weaknesses and monitor their progress. The strongest individuals are challenged to surpass their own achievements, while those having difficulty are invited to demonstrate what they can do and show the efforts they make to improve. No two portfolios are alike, since each one is a direct expression of a particular student.

4. Q: In portfolio assessment students are given credit for effort and improvement. Won't that unrealistically inflate the grades of these weak students who try hard?

A: Indeed, one of the strengths of portfolio assessment is that it enables students to be rewarded for efforts and progress, which traditional tests cannot do. However, this does not mean that a student who invests a lot of effort but performs poorly will get a high grade, since investment will be only one of the criteria for assessing the various tasks. These tasks will be assessed mainly on criteria relevant to their performance (for example, by rating scales). Weaker students know that they will always get credit for improvement, while remaining aware of the standards they are aiming for in the long run.

5. Q: How will portfolio assessment affect what I do in the classroom?

A: While tests are very different from activities in class, and thus may affect students' performance on language tasks, portfolio assessment helps you assess what students can actually do under normal conditions. In this way, your teaching can proceed as usual, though many of the activities students do in class will find their way into the portfolio, after undergoing revision. However, we will need to devote class time to such worthwhile activities as developing students' awareness of their own learning and improving their revision skills, in the context of whatever teaching content we have chosen.

6. Q: How do I know if I am doing a good job and that my students are progressing?

A: You will see progress from a first draft to a revised version, and from comparing the level of earlier pieces of work to later ones. Moreover, your students will show you evidence, in their comment cards, cover letters and the tutoring sessions we conduct with them, of the progress they are making, in a way that will give us much satisfaction.

7. Q: How will we find time for student-teacher conferences?

A: Try to integrate short talks with individual students into your working routine. It is impossible if you work only frontally, but is more feasible if you frequently activate the students individually and in pairs or groups for part of the lesson or during their free time to have "learning conversations.

8. Q: Must we correct all the items in each portfolio?

A: If we mean correcting all the mistakes, then the answer is a definite NO. Each assignment has its specific goals, and accuracy (i.e. speaking/writing without mistakes) may not be one of them, as in book tasks or writing journal entries. In such cases, assessment may focus on whether the task was completed and on criteria such as the student's investment or the insight shown in their response. In cases where accuracy is a goal, there are several possibilities:

- a. selective checking by the teacher (comments which guide the student towards improving specific aspects of the piece of work or correcting one type of mistake). As well as cutting down on the teacher's time spent checking, this is more effective than bombarding the student with comments on many types of mistakes at the same time;
- b. coded marking (for example. *sp* for spelling, *gr* for grammar) to guide students to correct their own mistakes;
- c. providing students with tools for assessing their own or their peer's work, which enable sharing of the responsibility for checking with the students themselves.

By the time you receive the student's portfolio and before it is being orally presented: you will have already given feedback on most of the entries, and will only have to monitor the improvement made.

9. Q: Now that I feel comfortable with portfolio assessment, should I give up tests?

A: Traditional tests will probably remain part of the educational system even when the benefits of alternatives in assessment have been firmly established. Therefore, we will probably use portfolio assessment alongside traditional tests,

and we may include tests as portfolio entries (corrected, of course). The English staff should draw up assessment plans (even better, do it with the students) which show the weighting of different assessment components (test, quiz, portfolio, project, extensive reading, homework, participation, etc.) and which instructional goals each tool assesses.

10. Q: How do I know the portfolio is the student's own work? A:

our best insurance against unreasonable "help" or plagiarism is to:

- a. require portfolio content to be based on work done in class;
- b. clarify that it is easy for the teacher to distinguish between a student's own work and an imported piece on an entirely different level;
- c. give clear guidelines to the students about the goals and criteria for excellence in a specific task. For example, downloading material, however colourful, from the Internet does not in itself demonstrate language skills; it is what the student can do with the source that counts (for example. summarize, compare different points of view, adapt etc.).

III-Answers to Students' Concerns about the Oral Tests

Our students have often expressed concerns about the Oral Test. Some indicate that they feel it is an arbitrary and unfair assessment tool. Most of the concerns arise from misunderstandings related to the oral Test, lack of experience from the teachers or inadequacy of the techniques used. This reported information is then passed on to others as a fact, ostensibly to explain why some students are not successful. Such "stories" have been circulating among language training students for many years. They serve to make learners mistrustful of the test, and put them at a disadvantage by making them even more nervous, thus interfering with their performance. Our goal is to make an

attempt to clear up these misunderstandings and provide facts that we hope will accurately reflect what takes place during an Oral Test.

Perception:

The assessor makes up his her mind about my level within the first few minutes of the Oral Test. Because first impressions count, I should try to sound as though I am not nervous and not make any mistakes in the warm-up in order to ensure that I will obtain a good mark.

Fact:

We do not think all the assessors make up their mind about your level within the first few minutes. They do get a preliminary idea about your level during the warm-up, but this is not rated. The final level is only determined once the whole test has been completed.

Perception:

The Oral Test is subjective. The assessor just talks to me for five/ ten minutes or so and arbitrarily decides what mark to give.

Fact:

The Oral Test is supposed to be conducted according to a specific structure and uses standardized techniques to elicit a sample of what you are capable of communicating in the target language. These standardized techniques provide for equivalency in testing from one candidate to the next, even though what is discussed varies among candidates. The system used to rate the speech samples uses highly defined criteria to determine when a candidate is allotted such or such a mark.

Perception:

Some teachers are more demanding than others.

Fact:

In this case, the test standards and techniques an equivalent form and content for all first year students, and measures are in place to ensure that these standards **are adhered** to by all teachers. It is true that before being accepted to administer the Oral Test, the teacher in charge must receive special intensive training and must demonstrate their ability to test and rate a wide range of candidates accurately and consistently. The objective nature of the test, the special training we expect our teachers to have, and the on-going quality control activities mean that each assessor must ensure that s/he or she administers and scores the test with the same method as the other assessors.

Perception:

The content of the test is difficult. Some of the questions will be hard to answer if I do not know anything about the topic. It is also important not to disagree with the assessor when expressing an opinion.

Fact:

During the Oral Test, the assessor is supposed to ask you questions about your work, your work experiences, or ask you to discuss other topics of interest that are related to your environment. The questions should be the ones that you have already covered in the course. If any of the questions asked during the test are sensitive for personal reasons, or if you do not know enough about a particular topic to be able to talk about it, you should inform the assessor and he or she will change the topic. Pedagogically, this should not have any effect on your test result. However, you do not have to be an expert on a topic to give an opinion. What is asked for is what you think and feel about a topic. During the test, the assessor should not put an exercise focus on the assessment of the factual content or ideas expressed (that is, how well and how much you know

about a specific topic), but rather measure how well you can communicate what you do know. In addition, the views expressed may be your own and need not be in agreement with what the assessor discusses.

Perception:

The test does not measure communication but style and knowledge, without considering my background and my personal style of speaking. I will have to speak quickly and talentfully. When we speak slowly we do not do well on the test.

Fact:

Although the Oral Test is highly structured, it should allow flexibility and should be tailored to the background of each candidate. The content of what is discussed should vary, depending on the topics to be discussed during the test. Assessors should be sensitive to the fact that candidates differ according to background, experience, and personality, as well as to ways of speaking and conveying information. For example, assessors should not be looking for rapid speech, but rather for speech that is smooth and without hesitations that may be distracting and make it difficult for the listener to follow. Assessors should be trained in all the techniques that are designed to fairly evaluate the linguistic performance of candidates who have different patterns of speech and communication and to adjust the interview to accommodate these differences.

Perception:

I have to use particular grammatical structures and sophisticated vocabulary during the interview and have little trace of an accent in order to do well on the test.

Fact:

Although it is a common misconception that knowledge of the grammatical elements of the target language, sophisticated vocabulary or

impeccable pronunciation is of paramount importance, the Oral Test in fact focuses on communicative proficiency rather than grammar, or pronunciation. While all of these are of course taken into consideration in determining the rating given to a candidate, what is considered is more the effect of the errors on the communication than the errors themselves. It is only when grammatical errors, lack of vocabulary or poor pronunciation interfere with the effectiveness of the communication that they are a problem. It is important that you respond to questions and participate in the conversation by concentrating on being clear about what you want to say (as you would in any conversation), and not focus on using a particular type of grammatical construction or highly sophisticated vocabulary in order to demonstrate that you have learned them. You should use whatever grammatical structure or vocabulary term is appropriate to convey your message and not try to use specific structures where they are not needed. For example, there are no specific or sophisticated linking terms that are expected, but appropriate linking terms will need to be used in order to convey a clear message.

Perception:

Assessors do not ask the "right" questions during the Oral Test. None of what I have prepared is asked about. Assessors only ask what they like.

Fact:

The Oral Test is not designed to evaluate memorized or prepared material or to allow you to demonstrate what you know in the second language, but rather to use what you know in the course of the interview. You demonstrate your proficiency in the second language by responding to the questions asked by the assessor; this ensures that you will be assessed according to how well you are able to use your second language.

Perception:

If the assessor knows the level that was assigned to me on a previous Oral Test, this will influence the level assigned on the next test. The assessor will use that information to help him/ her arrive at a decision for the rating.

Fact:

The rating of an Oral Test can only be determined by the sample provided during the administration of the test. There is no need for the assessor to take into consideration previous levels assigned, because the final rating assigned is representative of your proficiency at the time of the test. This level can be higher, the same or lower than a level you may have previously received.

IV- Advice and Action Plan for Students

We understand that many of our students face some difficulties in managing their time in the campus or in the faculty. To overcome the daily encountered difficulties, reduce their anxiety and increase the possibility to succeed, we recommend our students the following:

- Keep a sense of balance in your life between university work and leisure, do not get over tired and find time to relax. Sometimes, relaxation exercises can be helpful.
- Remember that there are many areas at the university and in your life which you are good at. Take some time to remember this, and build up your confidence. Think about all the successful and clever people who stammered and how they got on.
- Enlist the support of individual teachers in planning for oral work, and try to work with (a) friend(s), with whom you feel comfortable, whenever possible.

- Get as much information as you can about the examination requirements, **and work hard at forward planning and preparation. Use a cassette recorder to help you practise, and when you feel more confident, try out your talking on class mates and friends.**

- **Since you are allowed to choose your own subject, find one in which you are interested and take some pictures or use other visual aids to illustrate it.**

- **If you are talking about a text, make sure you really know it well.**

- **Plan out your talk in writing first, and then make a list of headings or key points on small cards. It is best to use this technique, rather than to read your talk from a script. Practise at home, recording your portfolio talk if you can and listening to your own presentation.**

- Remember that many students may feel like you do, and that some of them will find it very difficult to speak fluently when they are being tested. Just remember that teachers do not expect first year students to be totally fluent and perfect speakers; what they want is to hear you communicate your ideas and prepare your work thoroughly, showing good research, and appropriate listening skills and body language.

- The teacher will be on your side and will be valuing what you say just as much as how you say it.

- Take time and contact our psychologist in the Faculty or a chat about any feelings you may have of embarrassment, and this may help you a lot. You will probably find that other students, who you think are quite fluent, may deliberately speak more slowly when they give a talk because it helps.

- Try to speak and listen to English as much as possible before taking the test. You can do this by listening to the radio, watching television, and speaking English with your teachers and friends.

- Arrive on time and speak English from the beginning. Arriving on time and speaking English as soon as you meet the assessor will help you adjust more quickly to the testing session.
- Do not be overtly worried about making mistakes. If you can not think of a certain word, use a simple substitute to explain the meaning and continue with the conversation. If you are aware that you are making mistakes and would feel better if you corrected them, go ahead and do so. However, remember that frequent corrections may disrupt the flow of the conversation.
- Do not be discouraged if parts of the test seem difficult. At various times, the assessor will use more complex questions to give you the opportunity to perform at your maximum level of proficiency. However, testing at this higher level of proficiency will not take the entire testing session.
- Answer questions as fully as possible. In order to give the assessor a sufficient sample to evaluate, expand on your answers by giving details, explaining points or developing your thoughts.

V-Advice to Teachers

- Before students take a test, it may be useful to review these tips so that they are ready to do their best.
- Help your students understand that testing is a normal university activity and that it is part of the teaching/learning process.
 - Tell your students the purpose of the test they will take, how long they will wait for test results.
 - Briefly explain normal testing procedures. Tell them when and where they will take the test, how long the test will take. Tell them to arrive early at the test site; it will help them remain calm. If you monitor the test, be sure all the students fully understand the test directions.

- Tell your students to carefully mark each answer. If they have to erase an answer, they must erase it thoroughly to register their intended responses.
- Remind your students that they will have to stay aware of the time throughout the entire test and, if they finish early, they should review their answers.
- Guessing can have a great impact on your students' results. Find out how the particular test is scored and if students are penalized for guessing.
- Tell your students not to spend too much time on any one question. They may find it helpful to skip the hardest questions and then go back to them later.
- Students should always estimate their answers before working them out to determine if their calculations are reasonable.
- After the test, other recommendations have to be taken into consideration.
- Immediately after the test, encourage your students to talk about their experience. Discussions may help them to vent any frustration they feel about the testing process.
- When students receive their test results, do not discuss an individual's test scores in front of the entire class, rather talk about the class' performance as a whole. You may want to discuss these questions: Was the class's performance above or below the performance of other students in the other group? In what areas did the class do best? In what areas did the class have difficulty? Were the class' results consistent with day-to-day performance?
- If you discuss test scores with students, set up tutoring sessions with them. Encourage them to interpret their own performance. Do they think the scores accurately assessed their abilities or knowledge? How do their scores compare with their classroom performance? Were their scores higher or lower than they anticipated?

- **If you** discuss test scores with individual students, be **careful of the words you use**. Students may have a natural curiosity or anxiety about test results; they could be easily misled by unguarded comments.

- **Remind students that they should not make major decisions about their futures based on the results of one test. They need to consider their entire university history, their grade records, and their activities and interests before they solidify their career goals.**

Conclusion

We recognise that when we are involved in the process of transmitting knowledge and willing to serve the community, we need to devote plenty of time, energy and patience. It is in this light that the recommendations included in this chapter are made. The recommendations and advice centre around the management and the elaboration of a portfolio. This new vision of assessment based on autonomy of learning, can be introduced gradually into our system of education. Positive results may be obtained, with the commitment and the involvement of both teachers and learners and all the leading pedagogical / administrative staff. We hope our suggestions will serve as answers to questions that some teachers and learners ask themselves, when faced to learning situations.

GENERAL CONCLUSION

This research work aims at casting some light on a thorny issue in our system of education related to testing in general and proposing some alternative assessment techniques for measuring the oral proficiency of the population of first year students involved in the LMD System.

In this thesis the principles of language testing, issues related to the types and aims of testing and the characteristics of different tests are reviewed in terms of the basic considerations in test design covering the concepts of validity, reliability and practicality. The status of the various types of validity, and how the concept of validity relates to those of practicality are examined as well from a deeper theoretical perspective. The different approaches to language testing and their limitations in terms of the criteria of validity, reliability and practicality are described in terms of the particular focus they represent. The study has also looks into the possible uses of self- assessment through a portfolio, not only as an alternative to formal and traditional tests but basically and above all as a tool for learner independence as well as highlighting the importance of self- assessment as a study skill and as a measure to bridge the gap between testing and teaching. As we have taken into consideration some possible problems inherent to the nature of self- evaluation as well as anticipated problems related to the resources required for the implementation of such innovation for example, timing and training of staff and students.

Based on what has been investigated theoretically, we came to the conclusion that different investigations and research projects have shown that in very recent years, there has been a growing interest in the application of testing procedures that are totally different from traditional forms of assessment. These authentic forms of measurement have become increasingly popular in the

foreign language classes. Such forms of assessment are more student centred in **that, in addition to being an assessment tool, they provide the students with instruments to be more involved in their learning journey, and give them better sense of control and responsibility of their own learning. This is illustrated in the portfolio assessment which relates to the communicative approach to language testing which treats language as an instrument of interaction among human beings in society. The Portfolio technique generates more student involvement and some student appeal because of the relevant content of the input material, the authentic situations and the utility of the skills to be acquired and used. It also provides appropriate oral interactional situations, sometimes by leaving some participants free to choose what they think is their best production, they would like to share with their teacher and classmates, sometimes by providing deliberately conflicting and speech provoking questions. This is based on the assumption that language is essentially used for oral transactions.**

A critical evaluation of the testing situation in the Department of English has shown that a certain number of communicative testing techniques used by the teachers do not help much to foster in the students the rapid interpretation of utterances, the oral expression of meaning which is meant to lead the learners to an independent and autonomous position wherein they are expected to interact both genuinely and orally to talk about a topic of interest of their own choice.

The study seeks to demonstrate that the portfolio as an alternative form of oral self assessment can offer to both learner and teacher (tutor) valuable insight into the individual's cognitive processes, rather than simply revealing an ability to memorize and repeat. If this focus on the process of learning begins early in the programme of study, it means that the benefits of understanding one's own learning agenda are felt from the outset. It also allows time for the student to further develop as an effective learner, over time, gaining proficiency in the

range of learning skills most appropriate to his / her own context. An investigation departs from the belief that promoting this new vision to assessment will engender a dynamic pedagogy where the students are actors of their training and the educational team is a support, a guide and adviser all along their academic training. When students are asked to rate their current abilities, the academic team (tutors) will be prepared and ready to suggest and provide them with the tools and criteria with which they will assess themselves.

Based on these considerations and beliefs, we hypothesized that we can help first year LMD students involved in this study to increase their motivation and personal involvement in the learning process if we introduce the strategy of assessing one's abilities through the Portfolio Assessment Project at different points in the language learning process, with a teacher (a tutor) who plays a very determinant role in incorporating these strategies as cornerstones in the student's learning process and environment.

After analyzing the students' needs, the results have revealed that a large number of the informants (both teachers and students) have expressed their readiness and will to adopt the portfolio as an alternative, authentic form of oral self assessment. At the end of the process both learners and teachers (tutors) admitted that their experience in undertaking the portfolio was really a positive experience which offered them valuable insight into the individual's cognitive processes, rather than simply revealing an ability to memorize and repeat. Finally, the majority of the informants were convinced that the proposed alternative meets better their needs and expectations. Empirical investigation has shown that the learners have become aware of the need to self-assess in order to monitor their learning progress. For example, through portfolio and self-assessment, students have learnt that their progress and grade do not depend solely on the teacher, but in a large part on themselves. We have realized that

the students' produced portfolios and their self-evaluations have helped the **learners track areas of strength and weakness and address problems before final exams are issued. Self-evaluation has also allowed teachers to see how students view their progress, leading to individualized instruction, to initiate them to gain a progressive autonomy** and self reliance, and negotiate the best direction to take **in their learning process. From this perspective, we have realized that the teachers' assessment and learners' self- assessment** have proved to be a good assessment alternative since we reported a high correlation between the two, which is a response to an approach to testing which seeks to empower teachers and learners in the management of the educational process.

This research departed from what Alderson (1986: 105) concludes his article entitled "Innovations in Language Testing?" with: "**Perhaps we should be looking for** and fostering, not only innovations in language testing, but also and importantly, innovations through language testing." It starts from the premise that evaluation should not be seen as a necessary evil but as an intrinsic, constructive element of the teaching/ learning process. It is in this light that the initial aims of this study can be said to have been treated and accomplished not in any masterly fashion but in a way which we hope will raise awareness in teachers, learners, test experts methodologists, test designers and materials writers. Finally, it is our hope that the findings of this study will provide a tangible contribution to testing pedagogy and encourage the teachers of English to adopt the proposed alternative to assess the students.

APPENDICES

APPENDIX I: The L.M.D System	324
APPENDIX II: The Students' Questionnaire	334
APPENDIX III: The Teachers' Questionnaire	348

APPENDIX 1

The LMD System

1 -The Department of English and the New National Educational Reforms

The Department of English, Faculty of Languages and Human Sciences, University of Béjaia was founded in September 1991 and during that academic year, only 20 students enrolled for a licence in English.

Year by year, the department attracted more and more students and the enrolments in this licence increased considerably, reaching during the academic year 2005 a total number of one thousand and one hundred students. (this large number ,represents the importance given by graduates of Bejaia to learning foreign languages) .The needs of the learners were not totally met through the syllabus proposed to them, mostly when the graduates are in the job market facing serious difficulties. We have realized that the degree they try for does not prepare them to be operational in different job institutions, except to a certain extent in the field of teaching English as a foreign language. Because of the various dysfunctions in our higher system of education, decisions have been made by specialists to overcome the shortcomings. This new proposed device of education based on international Norms, is referred to as Licence, Master, Doctorate (LMD)

Because of the dysfunctions which appeared in the previous system of high education, specialists in charge of the file of reforms, realized that it is urgent and necessary to get the Algerian University out of the crisis in which it lives by equipping it, as soon as possible, with modern teaching, scientific,

human, material and structural means which will enable it to respond to the needs and expectancies of our society which is integrated into the international system of Higher Education. In order to deal with the various dysfunctions and difficulties currently met, as well on the level of management as on the level of the performances and output of the Algerian University, decisions have been made to implement a global and deep policy and reform of the Higher Educational System. This reform, while assuring the public character of high education, should reaffirm the essential principles which underlie the vision of the missions reserved for the Algerian university, namely:

- To ensure a high quality education by taking charge of the legitimate social need about having access to the higher education,
- To achieve a real osmosis with the socioeconomical environment by developing all possible interactions between the university and the world,
- To develop mechanisms of continuous adaptation to update professions,
- To strengthen its cultural mission by the promotion of universal values which are expressed by the academic mind, notably those of tolerance and respect of the other,
- To be opened more to the worldwide evolution, particularly the one of sciences and technologies,
- To encourage and to vary the international cooperation according to the most suitable shapes,
- To establish bases of good governance founded on involvement and dialogue.

In order to take up these challenges, some of the Algerian universities (university of Béjaia is one of them) have engaged deep reforms concerning the Higher Education based on:

- The offer of a high quality education for a better professional insertion,
- Education for all, all along life,

- Autonomy of universities,
 - and the opening of the university to the world.
-

2- Architecture of the New LMD System

The LMD system of teaching is organized towards an architecture articulated in three stages of education, each one conferring an academic diploma:

- Baccalaureate + 3 years, leading to the diploma of Licence,
- Baccalaureate +5 years, leading to the diploma of Master,
- Baccalaureate +8 years, leading to the diploma of Doctorate. In each of these stages, courses are organized in units of teaching regrouped in semesters of study.

The unit of teaching possesses the quality to be capitalizable and transferable, which means that its acquisition is definitive and that it could be used in another course of study.

This capitalization and this possibility of transfer allow the opening of bridges between several courses of education and lead to a mobility of students who will have a diversity of choices and the opportunity to follow a desired and individualized academic course.

The progression of studies is semestrial.

The courses Licence and Master have as main objective the acquisition of knowledge and the know-how driving to a professionalisation at two levels of qualifications. This system offers a general diagram that allows a progressive and adapted orientation through a suitable organization of teachings and profiles of education.

2.1-The "Licence"

This stage of education is organized in two landings. A basic and pluridisciplinary education lasting two to four semesters centred on the acquisition of fundamental principles related to the disciplines of the prepared diploma as well as to the initiation to the methodology of the academic life and its discovery. This landing is followed by a more specific education, offering two options:

- An «academic» option sanctioned by a diploma of "Licence" which allows the direct access to longer and more specialized higher education. This possibility is granted according to the required aptitudes, to the obtained results and to the established criteria of access.
- A "professionalising"and qualifying option sanctioned by a diploma of "Licence" which will allow a direct insertion into the labour world. The programs are defined in close collaboration with the using sector. This option requires, to be efficient, a multiplication of the offer of professional "Licences" corresponding to identified opportunities and should have a double effect. On one hand, to develop the diploma of "Licence" on the labour market as being that of the specialized professionals having sufficient bases to be able to evolve later on. It corresponds from this point of view, in the majority of the professions knowing fast transformations, to a required level requested by the professionals. In the other hand it should give a prospect to students who do not find a place in the stage of baccalaureate +5 years.

2.2-The "Master"

This period of education lasts two years. It is accessible to every student holding an academic licence meeting the conditions of access. It does not

exclude, therefore, students holding a professionalizing licence who will be able to return to the university after a passage in the professional life. This course prepares students to two quite distinct vocations:

- A professional vocation characterized by the acquisition of a greater specialization in a disciplinary field which will give access to levels of higher performance and competence. The orientation in this way remains always professional (Professional Master).
- Researcher vocation characterized by a preparation to the scientific research that predestines to research activities in the economic or in the academic sectors (Master of research).

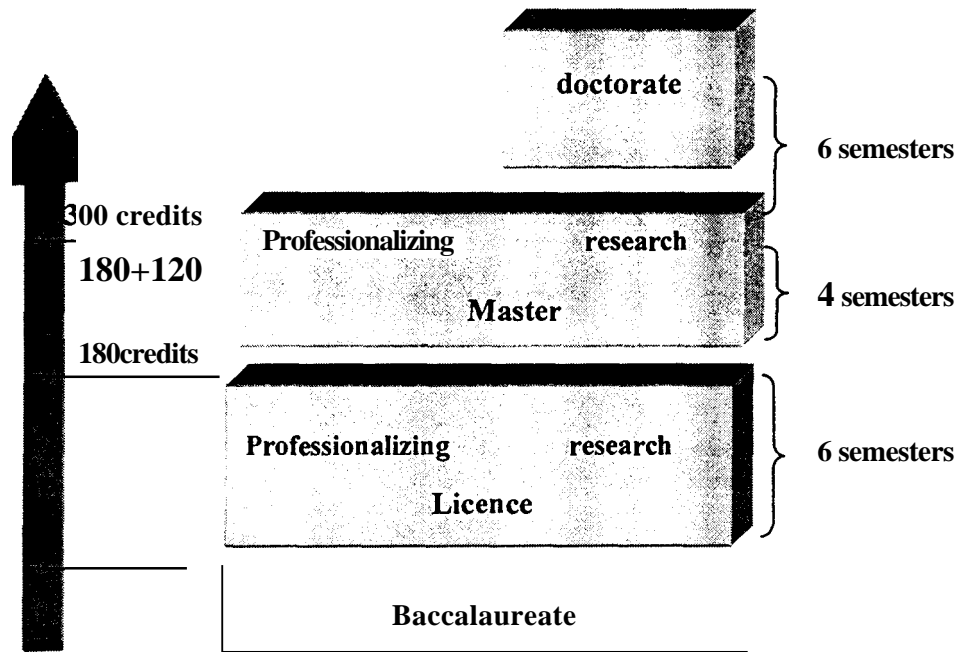
The Master constitutes the stamp of performances of the academic establishment.

2.3-The "Doctorate"

Facing the enormous development of knowledge and the more and more fine specializations and the more and more applied character of research, the doctoral training (of a minimal length of 6 semesters) should assure at a time: A deepening of knowledge in the field of specialism,

- And an education by and for research (development of abilities for research, sense of group-working ...). This education is sanctioned by a diploma of doctorate after submitting a thesis.

General diagram of the LMD system



3- Characteristics of the LMD System

The LMD system is based on a vision of a more coherent offer of education. These offers are presented by fields and are organized as standard courses. This system is also characterised by the possibility of mobility of both teachers and learners, and a large flexibility for students who are willing to decide, and become actors of both their learning itinerary and future career.

4-Training Areas

The new offer of education is organized inside great fields. A field is a regrouping of several disciplines in a coherent whole from the point of view of the Professional outlets to which it leads. For example:

Social and Human Sciences,(Languages,Arts,Sociology,Psychology
History. Philosophy)

Economic, Commercial, Financial and Management Sciences,

Exact sciences and Technologies (Mathematics/informatics, Physics,
Chemistry, Engineer Sciences)

- Health Sciences (Medicine, Pharmacy, Veterinary) etc.

In every domain a certain number of standard courses are defined which can correspond to specialties or options.

5-Standard Courses

A diploming standard course is a coherent combination of teaching units according to modes defined by the educational committees, according to an objective. It is validated by the tutelage. The various offered courses can integrate transdisciplinary, pluridisciplinary and professionalizing approaches, they allow, on one hand the student's progressive orientation according to his professional or personal project and taking into account, on the other hand, the diversity of publics and their needs and motivations.

6-Teaching Units

Studies are organized towards Teaching Units (TU) that are wholes of teachings (subjects) articulated in a coherent educational manner and according

to logic of progression in order to acquire identified competences. These units are taught by semester.

There are three Teaching Units categories:

6.1-The Fundamental Teaching Unit: It gathers the basic school subjects necessary to the continuation of the studies in the chosen branch.

6.2-The Discovery Teaching Unit: it regroups the matters of teaching that allows the student to widen the horizon of his knowledge and to open to him other perspectives in case of reorientation, thanks, notably; to interdisciplinary that characterizes its conception.

6.3-Transversal Teaching Unit: it regroups the subjects of teaching (foreign languages, informatics, technology of information and communication, humanities...) offering some necessary tools to the acquirement of a general culture and methodological techniques giving professional faculties of integration and adaptation into an environment in perpetual movement.

7-Credits

The credit is a teaching unit acquirement measure. Thus, every teaching unit has a determined value in terms of credits. This value, expressed by a number, is defined notably according to work likely to be done by the student (teachings, personal work, dissertation, project, training...). The definitive acquirements in terms of knowledge and competence are represented by credits. The credit constitutes a unit of value of one same ladder of reference shared by the whole academic establishments. It is, so, capitalizable and transferable. A common reference allows us to define the value in terms of credits of the whole

diplomas. This one is fixed to 180 credits for the Licence and 300 (180 + 120) for the Master. So the semester is valued to 30 credits.

8-Assessment and Progression

The student can get his diploma:

- by acquisition of each Teaching Unit
- by application of compensation modes between teaching Units

The assessment of teachings is done by continuous control of knowledge.

9-The Training "all along life"

The fast evolution of sciences and technologies requires a permanent updating of knowledge and competence level allowed only by a flexible system. The LMD system offers to all, and at all stages of life, whatever is the reached level and the expressed motivations, the opportunity to undertake or perfect an education a training, during or after a professional stay.

The same diploma can be acquired by different manners: initial education, continuous education, education in training and/or validation of the experience acquisitions.

10-The new Vision of the Training Policy

The LMD system has a new vision of the academic training centred on:

- an autonomy of the university on the basis of good governance,
- the development of a university project within the local, regional and national preoccupations on the economic and scientific levels and on social and cultural ones,

- offers of varied education and training, organized with a close collaboration of the economic sector,
- a dynamic pedagogy where the student is an actor of his training and the educational team is a support, a guide and adviser which come with him all along his learning process,
- a permanent and continuous assessment of establishments and teachings,
- mobilization and adherence of all the academic family.

In one word, the new reforms proposed if implemented, should create a vital and modern university that would prepare citizens, willing to serve their country, a University which is open to the world and holding the line with its environment.

APPENDIX II

THE STUDENTS' QUESTIONNAIRE

Dear student,

I would very much appreciate if you can give some of your time to answer this questionnaire which is part of a research work that aims to propose an adequate battery of tests which could be used in the evaluation of the students' speaking ability in the Department of English, Faculty of Languages and Human Sciences, university of Béjaia.

Your answers will be valuable for the completion of this work, and will be treated with great confidence.

Please, tick (J) the appropriate box (es) or give full answers where necessary.

May I thank you very much in advance for your cooperation.

Mr. Salah. DERRADJI

Department of Languages
Faculty of Letters and Languages

University Mantouri, Constantine

I - GENERAL QUESTIONS

1-How would you rate your oral proficiency in spoken English?

a-Very good

b-Good

c-Average

d-Poor

e-Very poor

2-Do you like to be tested?

Yes

No

3-Please, explain why?

.....

.....

4-Do you feel nervous before the test?

Yes

No

5-If “Yes”, please, explain.

6-Do you feel nervous during the test?

Yes

No

7-If “Yes”, please, explain.

II- PRINCIPLES OF TESTING

8- Is the teacher the sole responsible of testing?

Yes

No

9-Please, explain why.

10-Do you consider testing as:

a-Checking your performance?

b-Diagnosing your weaknesses?

c-Teaching?

e-Other: Please, specify

11-Do you fully exploit your testing sessions to enhance your oral performance?

Yes

No

12- If "No", please, explain why.

13-Are you regularly tested?

Yes

No

14-Do you like being regularly tested?

Yes

No

15-If “Yes”, please, explain why

III- BASIC CONSIDERATIONS IN TESTING

16- Are the tests you are given related to the topics covered
in the course?

Yes

No

17-If “Yes”, is it:

a -Always

b -Sometimes

c-Rarely

18- If “No”, please, explain why

19- Do the results obtained in the tests reflect your true level?

Yes

No

20-If "Yes", is it:

a -Always

b -Sometimes

c-Rarely

21-If "No", are you given:

a-Another test

b-Remedial work

c-Other: Please, specify:

...

..

22-If another teacher had conducted the test, would you have done better?

Yes

No

27- If “Yes”, is it:

a - Always

b - Sometimes

c- Rarely

28- Is the time allotted for the test adequate?

Yes

No

29-If “No”, what do you suggest?

IV-THE MARKING SCHEME

30-Are you informed of the marking scheme used by your teacher?

Yes

No

31-If "Yes", does it include?

a- Meaning

b- Grammar

c- Pronunciation

d- Stress and intonation

e- Fluency

f- Other: please, specify:

..

32-Does the marking scheme indicate clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable?

Yes

NO

33-Does it refer to your non linguistic features, such as attitude, gestures, way of dressing?

Yes

No

34- If "Yes" is it:

a - Always

b - Sometimes

c- Rarely

V-THE ORAL TEST

35-Which testing techniques are used to test your oral proficiency?

a-Multiple Choice Questions

b-Oral Interview

c- Group work

d-Role play

e-Discussion

f-Other: Please, specify:

36-Are these testing techniques adequate to test your oral proficiency?

a-Multiple Choice Questions	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
b-Oral Interview	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
c- Group work	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
d-Role play	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>
e-Discussion	Yes	<input type="checkbox"/>
	No	<input type="checkbox"/>

37- If "No", Please, explain why.

38-Do you understand the test directions in:

- | | | |
|-----------------------------|-----|--------------------------|
| a-Multiple Choice Questions | Yes | <input type="checkbox"/> |
| | No | <input type="checkbox"/> |
| b-Oral Interview | Yes | <input type="checkbox"/> |
| | No | <input type="checkbox"/> |
| c- Group work | Yes | <input type="checkbox"/> |
| | No | <input type="checkbox"/> |
| d-Role play | Yes | <input type="checkbox"/> |
| | No | <input type="checkbox"/> |
| e-Discussion | Yes | <input type="checkbox"/> |
| | No | <input type="checkbox"/> |

39- If your answer is "No", Please explain why.

40-Have you heard of self-assessment?

Yes

No

41-If "Yes", are you able to evaluate your oral proficiency?

Yes

No

42-Are you given the opportunity to evaluate your oral proficiency?

Yes

No

43-If is "Yes" is it through:

a-Assessing your own home or class work.

b-. Assessing your peers' home or class work

c-Other, Please, specify

VI-FURTHER SUGGESTIONS

44. Would you please add any suggestion or recommendation.

APPENDIX III

THE TEACHERS' QUESTIONNAIRE

Dear Teacher,

This questionnaire is part of a research work designed to devise a battery of tests which could be used in the evaluation of the students' speaking ability in the Department of English, Faculty of Languages and Human Sciences (University of Bejaia).

Your answers will be of great value as they will help the researcher to identify the problems faced by this population of learners, and the data will be treated with great confidence. Please tick ('J) the appropriate box(es) or make full statement whenever necessary.

May I thank you in advance for your collaboration.

Mr. Salah. DERRADJI

Department of Languages
Faculty of Letters and Languages
University Mantouri, Constantine

I-GENERAL QUESTIONS

1-How long have you been teaching?

2-How long have you been teaching oral expression?

3-How long have you been teaching oral expression in the first year?

4-Do you have any special training to teach oral expression?

Yes

No

II- PRINCIPLES OF TESTING

5-Is the teacher the sole responsible for testing?

Yes

No

6-Please, explain why.

7 - Testing is:

a-Checking student's performance.

b-Diagnosing students' weaknesses.

c-Teaching.

e-Other: Please, specify:

..

..

8-Do you fully exploit your testing sessions to enhance your students' oral Performance?

Yes

No

9- If “Yes”, please, explain how.

10- If “No”, please explain why?

11-Do you regularly test your students?

YES

NO

12-if “Yes”, please, say why?

III-THE TESTING PROCEDURE

13-Do you write your own tests?

Yes

No

14 -If "No", where do you choose them from?

a - A text book.

b- A colleague provides them.

c-Other: Please, specify:

.

.

15-Do you administer and conduct your oral examinations alone?

Yes

No

16 -Please, say why

.....

.....

IV-BASIC CONSIDERATIONS IN TESTING

17- Are the tests you conduct related to the topics you covered in the course?

-Yes

-No

18 -If “Yes”, is it:

a-Always

b-Sometimes

c- Rarely

19-If “No”, Please, say why

...

..

20-Do the results obtained from your tests reflect your students’ true level?

Yes

No

21-If “yes”, is it:

a- Always

b- Sometimes

c- Rarely

22-If “No”, do you give:

a- Another test

b- a Remedial work

c-Other: please, specify:

23-Are the test activities set at an appropriate level of difficulty?

Yes

No

24- If “Yes”, is it:

a -Always

b -Sometimes

c -Rarely

25-Does the test discriminate adequately between the performances of candidates at different levels of attainment?

Yes

No

26 - If "Yes", is it:

a-Always

b -Sometimes

c -Rarely

27- Are the abilities being rewarded those which the activities are designed to assess?

Yes

No

V- THE MARKING SCHEME

28- Does the marking scheme anticipate responses of a kind that candidates are likely to make?

Yes

No

29- Does the marking scheme indicate clearly the marks to be awarded for different parts of a question or the relative weighting of criteria that might be applicable?

Yes

No

30- Does the marking scheme allow for possible alternative answers?

Yes

No

31- Does the marking scheme, by specifying performance criteria, reduce as far as possible the element of subjective judgement that the examiner has to exercise in evaluating candidates' answers?

Yes

No

32- Can the marking scheme be easily interpreted by a number of different examiners in a way which will ensure that all mark to the same standard?

Yes

No

33-Which language aspects do you test? Specify the order of importance (1 for the most important one to 5 the least important one.)

a- Meaning

b- Grammar

c-Pronunciation

d-Stress and intonation

e-Fluency

f-Other: please, specify :

34-Do you evaluate non linguistic features, such as attitude, gestures, way of dressing?

35-If “Yes”, is it

a-Always

b -Sometimes

c -Rarely

36- If "No", Please, explain why.

VI- ORAL TESTS

37- Which testing techniques do you use to evaluate your students' oral proficiency?

a-Multiple Choice Questions

b-Oral Interview

c- Group work

d-Role play

e-Discussion

f-Other: Please, specify:

38-Have you heard of selfassessment?

Yes

No

39-If "Yes", do you think your students are able to evaluate themselves?

Yes

No

40-Please, explain why.

41-Do you give opportunities to your students to evaluate their oral proficiency?

Yes

No

42-If "Yes", is it through:

a- Assessing their own home or class work

b-Assessing their peers-home or classroom work

c-Other, Please, specify

43- Have you ever realized that the tests you administer to your students should take another form?

-Yes

-No

44- If "Yes", which ones?

45- What form should they take instead?

VI- FURTHER SUGGESTIONS

46- Would you please add any suggestion or recommendation?

Glossary of Important Testing Terms

Cohen, A.D. (1980), Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999)

The following terms and their definitions should prove useful in helping to understand the information provided in this research work.

Achievement test: measures what a learner knows from what he/she has been taught; this type of test is typically given by the teacher at a particular time throughout the course covering a certain amount of material.

Alternative assessment: refers to a non-conventional way of evaluating what students know and can do with the language; it is informal and usually administered in class; examples of this type of assessment include self-assessment and portfolio assessment.

Analytical scale: a type of rating scale that requires teachers to allot separate ratings for the different components of language ability i.e. content, grammar, vocabulary etc.

Artifact: (sometimes called students' evidence) It can include drawings, photos, video or audio tapes, writing or other work samples, computer disks, a research work done individually or in pairs in a library or collected from internet or from whatever source.

Authenticity: refers to evaluation based mainly on real-life experiences; students show what they have learned by performing tasks similar to those required in real-life contexts.

Autonomy: refers to developing strategies of Self and independent learning.

Banding scale: a type of holistic scales that measures language competence via descriptors of language ability; an example of this is the IELTS bands from UCLES.

Cloze procedure: a test construction procedure that involves deleting words on a systematic basis and replacing the deletions with blanks which the student must fill in.

Computer-based testing (CBT) is programmed and then administered to students on a computer; question formats are frequently objective, discrete-point items; these tests are subsequently scored electronically.

Computer-adaptive testing (CAT) presents language items to the learner via a computer; subsequent questions on the exam are "adapted", based on a student's response(s) to a previous question(s).

Content validity: this type of validity refers to testing what you teach; i.e. testing content covered in some way in the course materials using formats that are familiar to the student.

Cornerstones of good testing practice: the guidelines of effective test writers; they include the concepts of validity, reliability, practicality, transparency, authenticity, security and washback.

Construct validity refers to the fit between the theoretical and methodological approaches used in a program and the assessment instruments administered.

Criterion Referenced Test (CRT) is a test developed and used to estimate how much of the content and skills covered in a specific content area have been acquired by the examinee. Performance is judged in relation to a set of criteria

rather than in comparison to the performance of other individuals tested with a norm-referenced test (NRT). It compares students' performance to **particular** outcomes or expectations.

Descriptive statistics describe the population taking the test; the most common descriptive statistics include mean, mode, median, standard deviation and range; they are also known as the measures of central tendency.

Diagnostic test: a type of formative evaluation diagnoses students' strengths and weaknesses vis a vis the course materials; students receive no grades on diagnostic instruments.

Discrete-point test: an objective test that measures the students' ability to answer questions on a particular aspect of language; discrete-point items are very popular with teachers because they are quick to write and easy to score. These items are usually multiple choice, true-false, or fill-in-the-blank and allow for only one correct answer.

Distractor: an alternative-response choice which is intended to attract pupils who do not know the right answer.

Face validity refers to the overall appearance of the test; it is the extent to which a test appeals to test takers.

Formative evaluation refers to tests that are designed to measure students' achievement of instructional objectives; these tests give feedback on the extent to which students have mastered the course materials; examples of this type of evaluation include achievement tests and mastery tests.

Holistic Score is the assignment of a single score that reflects an overall impression of performance on a measure. Scores are defined by prescribed

descriptions of the levels of performance, examples of benchmarks at each level, or by scoring rubrics, an example of this is the scoring used with the TOEFL Test of Written English (TWE).

Integrative testing: goes beyond discrete-point test items and contextualizes language ability.

Inter-rater reliability attempts to standardize the consistency of marks between raters; it is established through rater training and calibration.

Item bank: a large bank or number of items measuring the same skill or competency; item banks are most frequently found in objective testing in particularly CBT and CAT.

Item Analysis: a procedure whereby test items and distractors are examined according to the level of difficulty of the item and the extent to which they discriminates between high-achieving and low-achieving students; results of item analyses are used in the upkeep and revision of item banks.

Language aptitude test is designed to discover whether learners possess a talent or a basic ability for learning a new language. It includes exercises in phonetic script, spelling clues, function of words in sentences and memorization of paired associate words.

Linguistic competence: the breadth of knowledge that the learner has regarding the elements of the language- pronunciation, vocabulary and structure.

Mean is known as the arithmetic average; to obtain the mean, the scores are added together and then divided by the number of students who took the test; the mean is a descriptive statistic.

Mode: the most frequently received score in a distribution.

Norm Referenced Test (NRT) is an instrument developed and used to estimate how the individuals being assessed compare to other individuals in terms of performance on the test. Individual performance is judged in comparison to other individuals tested, rather than against a set of criteria (criterion referenced test) or in a broad knowledge area (domain referenced test). It measures language ability against a standard or "norm" performance of a group; standardized tests like the TOEFL are norm-referenced tests because they are normed through prior administrations to large numbers of students.

Objective test: can be scored based solely on an answer key; it requires no expert judgment on the part of the scorer.

Parallel tests: multiple versions of a test; they are written with test security in mind; they share the same framework, but the exact items differ.

Performance-based test: requires students to show what they can do with the language as opposed to what they know about the language; they are often referred to as task-based.

Piloting: a common practice among language testers; piloting is a practice whereby an item or a format is administered to a small random or representative selection of the population to be tested; information from piloting is commonly used to revise items and improve them; also known as field testing.

Placement test: is administered to incoming students in order to place or put them in the correct ability level; content on placement tests is specific to a given curriculum; placement tests are most successfully produced in-house.

Portfolio assessment: one type of alternative assessment; portfolios are a representative collection of a student's work throughout an extended period of

time; the aim is to document the student's progress in language learning via the completion of such tasks as reports, projects, artwork, and essays.

Practicality: one of the cornerstones of good testing practice; practicality refers to the practical issues that teachers and administrators must keep in mind when developing and administering tests; examples include time and available resources.

Proficiency test: is not specific to a particular curriculum; it assesses a student's general ability level in the language as compared to all other students who study that language. An example is the TOEFL.

Quiz: a short test of class material, generally informal in nature.

Range: one of the descriptive statistics or measures of central tendency; the range or min/max is the lowest and highest score in a distribution.

Rating scale: instruments that are used for the evaluation of writing and speaking; they are either analytical or holistic.

Raw score: the score obtained directly as a result of totalling all the items answered correctly on a test.

Reliability: one of the cornerstones of good testing practice; it is the degree to which a test or assessment consistently measures whatever it measures. It is expressed numerically, usually as a correlation coefficient. There are several different types of reliability including:

-**Alternate Forms Reliability** is sometimes referred to as parallel forms or equivalent forms reliability. Alternate forms of a test are test forms to measure the same content area using items that are different yet equivalent. This type of reliability is conducted by correlating the scores on two different forms of the same test.

-Intra Rater Reliability is the degree to which a test yields consistent results over different administrations with the same individual performing at the same level by the same assessor (intra-rater).

-Inter Rater Reliability is the degree to which an instrument yields the same results for the same individual at the same time with more than one assessor (inter-rater).

-Internal Consistency Reliability is sometimes called split half reliability and is the degree to which specific observations or items consistently measure the same attribute. It is measured in a variety of ways, including Kuder Richardson 20 or 21, Coefficient Alpha, Cronbach's Alpha, and Spearman Brown Prophecy Formula. These methods yield a correlation coefficient that measures the degree of relationship between test items.

Rubric is sometimes referred to as a scoring rubric and is a set of rules, guidelines, or benchmarks at different levels of performance, or prescribed descriptors for use in quantifying measures of attributes and performance. Rubrics can be holistic, analytic or primary trait depending upon how discretely the defined behaviour or performance is to be rated.

Scoring a test: determining the number of points that each item or procedure is to receive, and then the value or weighting of these points with respect to the total test or quiz.

Self-assessment asks students to judge their own ability level in a language; one it is type of alternative assessment.

Specifications: a document that states what the test should be used for and who is it aimed at; test specifications usually contain all instructions, examples of test formats/items, weighting information and pass/fail criteria.

Standardized test: A carefully designed test that has undergone long experimentation and research. It is often administered to large groups of pupils from different types of schools. It measures language ability against a norm or standard.

Subjective test requires knowledge of the content area being tested; a subjective test frequently depends on impression and opinion at the time of the scoring.

Summative evaluation: refers to a test that is given at the end of a course or course segment; the aim of summative evaluation is to give the student a grade that represents his/her mastery of the course content.

Tutoring refers to the short meetings that teachers can have individually or in small groups with their students, in which progress is discussed and goals are set for a future meeting.

Validity: one of the cornerstones of good testing practice; refers to the degree to which a test measures what it is supposed to measure. A test is not valid per se; it is valid for a particular purpose and for a particular group. Validity evidence can come from different sources such as theory, research or statistical analyses. There are different kinds of validity including:

-Content Validity is the degree to which a test measures an intended content area. Content validity is usually determined by expert judgement of the appropriateness of the items to measure the specified content area.

-Construct Validity is the degree to which a test measures an independent hypothetical construct. A construct is an intangible, unobserved trait such as intelligence which explains behaviour. Validating a test of a construct involves testing hypotheses deduced from a theory concerning the construct.

-Concurrent Validity is the degree to which the scores on a test are related to the scores on another already established test administered at the same time, or to some other valid criterion available at the same time. The relationship method of determining concurrent validity involves determining the relationship between scores on the test and scores on some other established test or criterion. The discrimination method of establishing concurrent validity involves determining whether test scores can be used to discriminate between persons who possess a certain characteristic and those who do not, or those who possess it to a greater degree. This type of validity is sometimes referred to as criterion-related validity.

-Predictive Validity is the degree to which a test can predict how well an individual will do in a future situation. It is determined by establishing the relationship between scores on the test and some measure of success in the situation of interest. The test that is used to predict success is referred to as the predictor and the behaviour that is predicted is the criterion.

Washback: one of the cornerstones of good testing practice; refers to the impact a test or testing program may have on the curriculum.

Weighting refers to the value that is placed on certain skills within the exam determined through prior administrations to large numbers of students.

BIBLIOGRAPHY

- Alderson, J.C. (1978a). A Study of the Cloze Procedure with Native and Non-Native Speakers of English. University of Edinburgh Ph. D. Thesis.
- Alderson, J.C. (1978b). The Effect of certain Methodological Variables on Cloze Test Performance and its Implication for the Use of the Cloze Procedure in E.F.L. Testing. Mimeo. 5th International Congress of Applied Linguistics, Montreal.
- Alderson, J.C. (1983). "Who needs jam?" in Huges, A. & D. Porter (eds)
- Alderson, J.C. (1986). "Innovations in Language Testing?" in portal, M. (ed)
- Alderson, J.C., (1993). Judgments in language testing. In D. Douglas & C. Chappell (eds.), a new decade of language testing. Alexandria, VA, USA
- Alderson, **J.C. Clapham, C., & Wall, D. (1995)**. Language test construction and evaluation. Cambridge: Cambridge University Press. Anastasi, **A. (1982)**. Psychological Testing. London: Macmillan.
- Bachman, L. (1990)**. Fundamental Consideration in Language Testing, Oxford University Press.
- Bachman, L. and A.S. Palmer. (1981)**. "A multitrait-multimethod investigation into the construct validity of six tests of speaking and reading" in palmer, A.S et al. (eds)
- Bachman, L. and A.S. Palmer. (1981)**. Basic Concerns in Test Validation. In Read (ed.) 1981a: 41-57.
- Bachman, L. and A.S. Palmer. (1981b)**. The construct validation of the F.S.I Oral Interview. Language Learning 31/1: 67/86.
- Bachman, L. and A.S. Palmer. (1984)**. **Some comments** on the terminology of language testing. In: Rivera, C. (ed.), Communicative

- Competence Approaches to Language proficiency Assessment: Research and Application. Clevedon: Multilingual Matters, 34-43. **Bachman, L. and A.S. Palmer (1996)**. Language testing in practice. Oxford:
- Bateson, D. (1994)**. Psychometric and philosophic problems in "authentic" assessment: Performance tasks and portfolios. *Alberta Journal of Educational Research*, 40 (2), p. 233-245
- Berkoff, N. A. (1985)**. Testing oral proficiency: A new approach. In Y. P. Lee, A. C. Y. Y. Fok, R. Lord & G. Low (eds.), *New directions in language testing* (pp. 93-99). England: Pergamon Press.
- Bloom, B.S.T.J. Hastings and G.F. Madaus (1971)**. *Handbook on Formative and Summative Evaluation of Student Learning*, New York: McGraw-Hill.
- Breen, M and C. Candlin (1980). "The essentials of a communicative curriculum in language teaching" in *Applied Linguistics* 1, 89-112.
- Breen, M.P et al (1997). *Communicative Materials Design: Some Basic Principles* In practical papers in English Language Education vol 2 university of Lancaster.
- Briggs, S. and C. MacDonald. (1978). *A Practical Approach to Testing, Speaking and Listening Skills*. *Forum* 16: 8-15.
- Brown, G., A. Anderson, R. Shillcock and G.Yule. (1984). *Teaching Talk Strategies for Production and Assessment*, Cambridge University Press.
- Brown, J. D. (1988)**. *Understanding Research n Second Language Learning*, Cambridge: Cambridge University Press.
- Brown, J. D. (1995)**. *The elements of language curriculum*. Boston: Heinle & Heinle.

- Brown, J. D. (1996).** Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents.
- Brown, J. D., & Hudson, T. (2002).** Criterion-referenced language testing. Cambridge: Cambridge University Press.
- Brown, S. Rust, C., Gibbs, G., (1994).** Strategies for Diversifying Assessment in Higher Education, Oxford, Oxford Centre for Staff Development.
- Campbell, D.T. and D.W. Fiske (1959).** Convergent and Discriminant Validation by the Multi-Trait/ Multi-Method Matrix, Psychological Bulletin 56: 81-105.
- Campbell, R. and R. Wales. (1970).** The Study of Language Acquisition. In Lyons J. (ed) New Horizons in Linguistics. Penguin: 242-260.
- Canale, M. (1983).** On some Dimensions of Language Proficiency, in Oller (ed.): 333-342.
- Canale, M. (1985).** "Proficiency-oriented Achievement Testing" paper presented in the Master series of the American Council on the teaching of foreign languages, at the Defence Language Institute in Monterrey, San Francisco and San Antonio.
- Canale, M. and Swain, M. (1980).** Theoretical bases of communicative approaches to second language teaching and testing. Journal of Applied Linguistics, 1, 1: 1-47.
- Canale, M. and M. Swain. (1980).** Theoretical Bases of Communicative Approaches to second Language Teaching and Testing, Applied Linguistics I: 1-47
- Canales, J. A. (1994).** Linking Language Assessment to Classroom Practices. In R. Rodriguez, N. Ramos, & J. A. Ruiz-Escalante (eds.) Compendium of Readings in Bilingual Education: Issues and Practices. Austin, TX: Texas Association for Bilingual Education.

- Candlin, C.N. (1986).** Explaining Communicative Competence: Limits on Testability, in Stansfield. (ed.): 38-57.
- Carroll, B.J. (1967).** "Foreign Language Proficiency Levels Attained by Language Majors near Graduation from College." *Foreign Language Annals* 1.1 (1967): 131-51.
- Carroll, B.J. (1978).** Specifications for an English Language Testing Service, London: British Council.
- Carroll, B.J. (1980).** *Testing Communicative Performance*, Oxford: Pergamon Press.
- Carroll, B.J. (1980b).** *Testing Communicative Performance: An Interim Study*. Oxford: Pergamon.
- Carroll, B. J., & Hall, P. J. (1985).** *Make your own language tests: A practical guide to writing language performance tests*. Oxford: Pergamon Press
- Chaplen, E.F. (1970a).** *The Identification of Non-Native Speakers of English Likely to Under-Achieve in University Courses Through Inadequate Command of the Language*. University of Manchester Ph. D. Thesis.
- Chaplen, E.F. (1970b).** *Oral Examinations*, in Centre for Information on Language Testing and Research, *Examining Modern Languages*, London: C.I.L.T. et Committee on Research and development in Modern Languages (C.I.L.T. Reports and Papers 4): 38-45.
- Cohen, A.D. (1980).** *Testing Language Ability in the classroom*. Newbury House, Rowley, Mass.
- Cohen, A.D. (1994).** *Assessing language ability in the classroom*. Boston: Heinle & Heinle.
- Cohen, A.D. (2001).** *Second language assessment*. In M. Celce-Murcia (ed.), *Teaching English as a second or foreign language* (3rd ed., pp. 515-534). Boston: Heinle & Heinle.

- Cronbach, L.J. (1971). Test Validation, in Thorndike (Ed.): 443-507.
- Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (ed.), Language proficiency and academic achievement. Avon, England: Multilingual Matters Ltd.
- Davies, A. (1965). Proficiency in English as a Second Language. University of Birmingham Ph. D. Thesis.
- Davies, A. (1977). "The construction of language tests" in Allen, J.A.B. & A. Davie (eds) The Edinburgh Course in Applied Linguistics, vol. 4: Testing and Experimental Methods, Oxford: Oxford University Press.
- Davies, A. (1978). Language Testing. Survey Article Parts I and II. Language Teaching and Linguistics Abstracts II 3/4/ 145-159 and 215-231.
- Davies, A. (1981a). A Review of Communicative Syllabus Design. T.E.S..O.L. Quarterly 15/3: 332-338.
- Davies, A. (1981b). Reaction to the Palmer and Bachman and the Vollmer Papers (2), in Alderson and Hughes (eds.): 182-186.
- Davies, A. (1983). The validity of Concurrent Validation. In Hughes, A. and Porter, D. (eds.), Current Developments in Language Testing. London: Department of Linguistic Science University.
- Davies, A., **Brown, A., Elder, C., Hill, K., Lumley, T., and McNamara, T.** (1999). Dictionary of language testing. Cambridge: Cambridge University Press.
- Dickinson, **L.** (1987). Self-instruction in Language Learning, Cambridge: Cambridge University Press. Ellis, G & B. Sinclair (1989) Learning to Learn English, Cambridge: Cambridge University Press.
- Dickson, **L.** (1981). Have you got Mr Bum the Baker problems and solutions in the use of Games. Role play and Simulations

- Francis, J.O (1981). The reliability of two methods of marking oral tests in modern language examinations, in *British Journal of Language Teaching* 19, 15-23
- Hall, C. (1993): The direct testing of oral skills in university foreign language teaching, in *IRAL* 31, 23-38
- Fromkin, V and R. Rodman. (1983).** An introduction to Language. New York. Genesee, F., and Upshur, J. A. (1996). Classroom-based evaluation in second language education. Cambridge: Cambridge University Press.
- Gimson. A.C. (1989).** An introduction to the pronunciation of English. Revised by Susan Ramsaran. London
- Graves, K. (2000).** Designing language courses: A guide for teachers. Boston: Heinle & Heinle.
- Guilford, J. P. (1965).** Fundamental statistics in Psychology and Education. 4th Edition. MCGRAW-Hill.
- Hawkey, R. (1982). An Investigation of Inter-Relationships Between Cognitive/Affective and Social Factors and Language Learning. A Longitudinal Study of 27 Overseas Students Using English in Connection with Their Training in the United Kingdom. University of London Ph. D. Thesis.
- Heaton, J. B. (1975).** Writing English Language Tests, London: Longman.
- Heaton, J. B. (1982).** Language Testing, Oxford: Modern English Publications.
- Heaton, J. B. (1990).** Classroom testing. New York: Longman.
- Henning, G. (1987).** A guide to language testing. Cambridge, Massachusetts: Newbury House.
- Holec, H. (1980). Autonomy and Foreign Language Learning, Strasbourg: Council of Europe.
- Holec, H. (1988). Autonomy and self- directed learning: present fields of application Strasbourg Council of Europe.
- Hughes, A. (1989). Testing for language teachers. Cambridge: Cambridge University Press.

- Hughes, R. (2002). Teaching and researching speaking. London: Pearson Education Limited.
- Hymes, D.H. (1972). On communicative competence, in pride and Holmes (eds)
- Ingram, D.E. (1974).** "Language Testing" in Allen, J.P.B. & S. P. Corder (eds) The_Edinburgh Course in Applied Linguistics: techniques in applied linguistics vol.4. London: Oxford University Press.
- Ingram, D.E. (1977).** Basic concepts in testing, in Allen & Davies (eds): 11-37.
- Jakobovits, L.A. (1970).** Foreign language learning: A psycholinguistic analysis of the issues. Newbury House.
- Kelly, R. (1978).** On the Construct Validation of Comprehension Tests: An Exercise in Applied Linguistics. University of Queensland Ph. D. Thesis.
- Kelly, R. (1980).** From innovation to adaptability: The changing perspective of curriculum development. In Galton, M. (ed.), Curriculum change: The lessons of a decade, 65-80. Leicester, England: Leicester University Press.
- Lado, R. (1961).** Language Testing: The construction and use of foreign language tests. London: Longman.
- LeBlanc, R. G. Painchaud (1985).** "Self-Assessment as a second language placement instrument" in TESOL Quarterly, vol. 19 n° 4 Dec. 1985.
- Madsen, H. S. (1983).** Techniques in testing. New York: Oxford University Press.
- Madsen, H. S., and Jones, R. L. (1981).** Classification of oral proficiency tests. In A. S. Palmer, P. J. M. Groot & G. A. Trosper (Eds.), The construct validation of tests of communicative competence (pp. 15-30). Washington, DC: TESOL.

- McLeod, N. (1983). Some techniques for involving learners in their own evaluation, unpublished notes, British Council Language Teaching Centre, Jakarta, Indonesia.
- Messick, S.A. (1980). "Test validity and the ethics of assessment", in *American Psychologist* 35: 1012-27.
- Milanovic, M. (1988). The construction and validation of a performance- based battery of English language progress tests, unpublished PhD dissertation, University of London.
- Moller, A.D. (1981a). Assessing Proficiency in English for Use in Further Study, in Read (ed.): 58-71.
- Moller, A.D. (1981b).** Reaction to the Morrow paper (2), in Alderson and Hughes (eds.): 38-44.
- Moller, A.D. (1982b).** A Study in the Validation of Proficiency Tests of English as Foreign Language. University of Edinburgh Ph. D. Thesis
- Morrow, K. (1979). "Communicative language Testing: revolution or evolution, In Brumfit, C. J. & K. Johnson (eds) *The Communicative Approach to Language Teaching*, Oxford: Oxford University Press.
- Morrow, K.E. (1986).** The Evaluation of Tests of Communicative Performance, IN Portal (ed.): 1-13.
- Mugglestone. P. (1977).** Role Play E.L.T Documents.: Games, Simulations and Role playing. London. British Council
- Munby, J.L. (1978).** *Communicative Syllabus Design*, Cambridge: Cambridge University Press. North, B. (1990) *Assessment in the Classroom*, paper presented at the 24 the IATEFL Conference, Dublin.
- Murphy, R.J.L. (1978).** Sex Differences in Objective Test Performance. Mimeo. Associated Examining Board Research Report RAC/56.

- Murphy, R.J.L. (1979).** Mode 1 Examining for the general Certificate of Education. A General Guide to some Principles and Practices. Mimeo. AEB. Guildford.
- Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998).** Designing second language performance assessments. Honolulu, HI: University of Hawaii Press.
- O'Malley, J. M., and Valdez Pierce, L. (1996).** Authentic assessment for English language learners: Practical approaches for teachers. New York: Addison Wesley.
- Oiler J.W. (1973).** Pragmatic language Testing. *Language Sciences* 28: 7-12
- Oiler J.W. (1979).** *Language Tests at Schools.* Longman.
- Oiler, J.W. Jr. and Damico, J.S. (1991).** Theoretical considerations in the assessment of LEP students. In E. Hamayan & J.S. Damico (Eds.), *Limiting bias in the assessment of bilingual students.* Austin: Pro-ed publications.
- Oskarsson, M. (1978).** *Approaches to Self-assessment in Foreign Language Learning,* Council of Europe, Oxford: Pergamon Institute of English
- Oskarsson, M. (1981).** Subjective and Objective Assessment of Foreign Language Performance, in Read (ed) 225-239.
- Oskarsson, M. (1984).** *Self Assessment of Foreign Language Skills: A survey of research and development work.* Strasbourg : Council of Europe
- Oskarsson, M. (1989).** " Self-assessment of language proficiency: rationale and applications", in *Language Testing*, vol. 6 nol June 1989 London: Edward Arnold. Oxford University Press.
- Palmer, A.S., **P.J.M. Groot, and G.A. Trosper. (Eds.) (1981a).** *The Construct Validation of Tests of Communicative Competence.* T.E.S.O.L.: Washington D.C.

Paulson, F.L., P.R. Paulson and C.A. Meyer. (1991). what makes a portfolio a a portfolio? Educational Leadership.

Pilliner, A.E.G. (1968). Subjective and Objective Testing. In A. Davies (ed.).
Language testing symposium: A psycholinguistic approach. (pp.
19-35). Oxford University Press.

Porter, D. (1983). Assessing communicative proficiency: the search for
validity, in Johnson and porter (eds) 1983: 189-203
Portfolio? Educational Leadership. February.

Potts P.J. (1985). "The role of evaluation in a communicative curriculum, and
some consequences for materials design", in Alderson, J.C. (ed)

Rea Dickens and Germaine. (1993). Evaluation.Oxford: Oxford University
press.

Rea, P.M. (1978). assessing language as communication .MALS. journal New
series no.3.

Rea, P.M. (1985). "Language testing and the communicative teaching curriculum"
, in Lee, Y.P. ET .AL. (EDS):15-32

Read, JAS. (ed)(1981a). Directions in language testing SEAMEO regional centre
.anthology series 9

Rechterich, R and Chancerel, J.L. (1987). Identifying the Needs of Adults.
Prentice Hall International.

Revel, J. (1979). Teaching Techniques for Communicative English. Macmillan

Richards, J. C. and Rodgers, T. (1986). Approaches and methods in language
teaching: A description and analysis. Cambridge: Cambridge
University Press.

Robinson, P. (1973). "Oral expression tests", in English Language Teaching, vol .
25 no 1, 2-3.

- Robinson, P. (1981). Role Playing and class participations. *ELT Journal* vol 3f N° 4 July 81
- Robinson, P.O (1981).E.S.P/ the Current Position Pergamon
- Savignon, SJ. (1972). teaching for communicative competence : a Research report . *audio visual language journal* 10/3: 153-162
- Skehan, p. (1984).Forthcoming .variability and language testing, in Ellis (ed) Skehan,P. (1988/9). "Language Testing: survey article parts 1 and 2", in *Language Teaching Abstracts*, vol.21, 4 and vol. 22, 1.(1989) *Individual Differences in Second-Language Learning*, London: Edward Arnold.
- Smith, K. (1989). Self-evaluation in the Foreign Language Classroom, paper presented sat the 23^d IATEFL Conference, Warwick
- Spolsky, b.(ed) (1968). Language testing: the problem of validation *TESOL Quarterly* 2:88-94
- Spolsky, b.(ed) (1976). language testing: art or science paper read at the4th international congress of applied linguistics. Stuttgart : hochschulverlag . Germany.
- Stevenson, D.k. (1985a).Authenticity, validity and a tea party language testing 2/1 :41/47
- Stevenson, D,k. (1985b). Pop validity and performance testing, in lee et at (eds):111-118
- Swan, M. (1985). "A critical look at the communicative approach (1)" in *ELT Journal* vol . 39/1 Jan 1985.
- Swain, M (1985). Large-scale communicative language testing: a case study, in lee et al (eds): 35-146
- Trump, J.L. (1967). "Influencing change at the secondary level" in Miller, R (ed) *Perspectives on Educational Change*, New York: Appleton Century-Crofts.

- Underhill, N. (1987).** Testing spoken language. Cambridge: Cambridge University Press.
- Valdés, G. and Figueroa, R. (1994).** Bilingualism and testing. A special case of bias. Norwood, NJ: Ablex Publishing Corporation.
- Von Elek, T. (1985).** "A test of Swedish as a second-Language: an experiment in Self-assessment", in Lee, Y.P. et. Al. (eds).
- Wall, D. (1982a).** Study of the predictive validity of the Michigan battery, with special to the communication. In culhane, klein-braley and Stevenson (eds): 156-169
- Weir, C.J. (1983).** The associated examining board's test in English for academic purposes: an exercise in content validation, in Hughes and porter (eds): 147-153
- Weir, C. J. (1990).** Communicative language testing. London: Prentice Hall International.
- Weir, C. J. (1995).** Understanding & developing language tests. New York: Phoenix ELT.
- Wesche MB (1987).** "Second Language performance testing: the Ontario test of ESL as an example", in Language Testing, vol. 4, 1 28-47
- Widdowson, H. G. (1978).** Teaching Language as Communication, Oxford: Oxford University Press.
- Widdowson, H.G. (1983).** Learning Purpose and Language Use, Oxford: Oxford University Press. xxxv N04. July 1981

II-LE RÉSUMÉ

La présente étude traite des questions en rapport avec les tests de langue et propose une batterie d'épreuves qui pourrait être utilisée dans un contexte d'évaluation des capacités orales chez les étudiants de la première année LMD. Nos Enquêtes et recherches ont montré que, durant les dernières années, un grand intérêt a été accordé à l'application des procédures d'évaluation entièrement différentes des formes traditionnelles. De nombreuses formes authentiques d'évaluation sont, de plus en plus, en vogue dans les classes d'enseignement de la langue étrangère.

Cette "évaluation authentique", est appelée "évaluation portfolio", où les apprenants devraient démontrer, plutôt que relater, ce qu'ils savent et peuvent faire. Les Contenus de "l'évaluation portfolio" (parfois appelée des "artifacts" ou des "preuves") peuvent inclure dessins, photos, vidéo ou cassettes audio., écrits ou autres échantillons de travail, disquettes, travaux de recherche faits individuellement ou en binômes dans une bibliothèque ou tirés d'Internet ou d'une quelconque source. Cette forme d'évaluation, outre son caractère d'outil d'évaluation, est beaucoup plus centrée sur l'étudiant, elle offre aux apprenants des instruments susceptibles de les impliquer dans leur classe d'étude, et leur donne un meilleur sens de contrôle et de responsabilité à l'égard de leur propre apprentissage.

Cette nouvelle forme, qui se veut une alternative d'évaluation des capacités de l'étudiant au niveau de l'expression orale, offre à l'apprenant ainsi qu'à l'enseignant (tuteur) une perspicacité précieuse dans les processus cognitifs de l'individu, elle ne permet pas simplement de révéler une capacité de mémoriser et de répéter. Elle offre, en outre, le temps suffisant à l'étudiant pour développer ses compétences de manière très efficace à travers le temps et

acquérir des capacités dans le rang des habiletés d'apprentissage les plus appropriées à son propre contexte. L'enseignant (tuteur) a aussi un rôle déterminant dans l'incorporation de ces stratégies comme pierres angulaires dans le processus d'apprentissage et dans l'environnement de l'apprenant.

Nous avons confirmé l'hypothèse que nos étudiants ont besoin d'une auto-évaluation afin de pouvoir diriger leur progression dans l'apprentissage. Par exemple, les étudiants peuvent apprendre à travers l'auto évaluation et le Portfolio, que leur progression ainsi que leur niveau ne dépendent pas uniquement de l'enseignant mais dans une grande partie, d'eux-mêmes. Appliqué régulièrement, le portfolio et l'auto évaluation sont à même d'aider les étudiants à prendre connaissance de leurs points de force et de faiblesse et de régler tous leurs problèmes avant les examens finaux. Par ailleurs L'auto évaluation permet aussi aux enseignants de voir comment les étudiants envisagent leur progression, les menant ainsi à s'auto instruire individuellement et à s'initier à un apprentissage autonome.

Cependant, nous croyons fort bien que les étudiants ne peuvent avancer dans leur carrière comme apprenants actifs s'ils ne sont pas capables d'évaluer leurs propres capacités et progression et tracer une trajectoire pour l'apprentissage futur. Les étudiants ont besoin qu'on leur apprenne explicitement les stratégies de l'auto évaluation et qu'on les aide à mettre en oeuvre ces stratégies. La plus grande valeur de l'auto évaluation et de l'évaluation portfolio réside dans le fait que les étudiants, durant leurs parcours de formation, deviennent contributeurs actifs dans le processus d'apprentissage et d'évaluation.

La promotion de cette nouvelle vision de l'évaluation est susceptible d'engendrer une dynamique pédagogie où l'étudiant constitue l'acteur principal dans sa formation et où l'équipe pédagogique représente un support, un guide et

un conseiller qui accompagne l'étudiant tout au long de sa formation académique. l'équipe académique (tuteurs) sera à disposition des apprenants pour leur suggérer et leur fournir les outils et critères qui serviront à leur auto évaluation.

III- ملخص

تتناول هذه الدراسة المسائل ذات الصلة باختبارات اللغة، و تقترح جملة من الامتحانات التي يمكن استعمالها في نطاق تقييم القدرات الشفوية لدى طلبة السنة الأولى ل.م.د. لقد بينت البحوث التي أجريناها أن اهتماما خاصا أسند خلال السنوات الأخيرة لإجراءات تقييم مختلفة تماما عن الأشكال التقليدية، أصبحت أشكال التقييم الأصلية أكثر فأكثر استعمالا في أقسام تعليم اللغة الأجنبية.

إن هذا التقييم الأصيل المسمى بتقييم "بورتوليو" يجعل المتعلمين يبرهنون بدل سرد ما يعرفونه و يستطيعون القيام به. إن مضامين تقييم "بورتوليو" (المسمى في بعض الأحيان أو بالحجج) بإمكانهم أن تحتوي رسوم، صور، فيديو أو أشرطة صوتية، كتابات وعينات أخرى من الأعمال كالأشغال المنجزة فرديا أو ثنائيا في مكتبة أو مستخرجة من الأنترنت و غيرها من المصادر. إن هذا الشكل من التقييم، زيادة عن كونه وسيلة تقييم، فهو أكثر تركيزا على الطالب، إذ يمنح للطلبة وسائل قادرة على إدماجهم في قسمهم الدراسي وتعطيهم أفضل إحساس بالتحكم والمسؤولية تجاه عملية تعلمهم.

إن هذا الشكل الجديد للتقييم الهادف إلى تقديم تقييم بديل لقدرات الطلبة على مستوى التعبير الشفوي، يمنح للمتعلم و المعلم (الولي) على حد سواء فعالية قيمة في عمليات الفرد المعرفية، فهي لا تؤدي إلى إبراز قدرة الذاكرة و التكرار فحسب، بل تتيح كذلك الوقت الكافي للطلبة لتنمية كفاءاته بطريقة ناجعة من خلال الزمن ومنه اكتساب قدرات في مصفّ مهارات التعلم الأكثر تناسبا مع سياقها. للمعلم (الولي) دور هام في إدماج هذه الإستراتيجيات كعناصر أساسية في عملية التعلم و محيط الطالب المتعلم.

لقد أكدنا الفرضية التي مؤداها أن الطلبة في حاجة إلى تقييم ذاتي من أجل التمكن من توجيه تقدمهم في التعلم. إن الطلبة، على سبيل المثال، بإمكانهم التعلم عن طريق التقييم الذاتي و البورتوليو ولا يرتبط تقدمهم و كذا مستواهم بالمعلم بحسب، وإنما يرتبط كذلك وإلى حد بعيد بأنفسهم. إن تطبيق البورتوليو والتقييم الذاتي يساعد الطلبة على التعرف على نقاط قوتهم وضعفهم وكذا حل مشكلاتهم قبل الامتحانات النهائية. من جهة

أخرى، يسمح التقييم الذاتي للمعلمين بإدراك كيفية توقع الطلبة لتقدمهم، مما يؤدي بهم فرديا إلى التعلم الذاتي و التعلم المستقل.

إلا أننا نعلم جازمين بأن الطلبة لا يتقدمون في مسارهم كمتعلمين نشطين (فاعلين) إذا لم يتمكنوا من تقييم قدراتهم الذاتية وتقدمهم ومن رسم معالم تعلمهم المستقبلي، فالطلبة في حاجة إلى تعلم استراتيجيات التقييم الذاتي وإلى مساعدتهم على تطبيق هذه الاستراتيجيات. إنّ أفضل قيمة للتقييم الذاتي وتقييم بورتفوليو تكمن في كون الطلبة يتحولون خلال مسارهم التكويني إلى مساهمين فاعلين في عملية التعلم والتقييم.

إنّ ترقية هذه النظرة الجديدة للتقييم قادرة على توليد ديناميكية بيداغوجية يشكل فيها الطالب الفاعل الأساسي في التكوين، ويمثل فيها الفريق البيداغوجي الوسيلة، الموجه والمستشار المرافق للطالب طوال تكوينه الأكاديمي، عندما يدعى الطلبة إلى تقييم قدراتهم الراهنة، يقوم الفريق الأكاديمي (الأولياء) باقتراح ووضع رهن تصرفهم الأدوات والمقاييس اللازمة لإجراء تقييم ذاتي.