

THE CONSTRUCTION AND VALIDATION OF A TEST IN ENGLISH FOR
TERTIARY EDUCATION WITH REFERENCE TO ADDIS ABABA UNIVERSITY

Teshome Demisse

A Thesis Submitted for the Degree of Doctor of Philosophy
in the Teaching of English as a Foreign Language (TEFL)
Department of Foreign Languages and Literature
Institute of Language Studies
Addis Ababa University

June 1995

ABSTRACT

This study explores the validity of a battery of tests of English for academic purposes. A goal-oriented, skills- and task-based test of English is developed and validated as a measure of the English language disposition of Freshman students at Addis Ababa University.

The test is based on analysis of the communicative language needs of the students (Morris, 1982; Haile Michael, 1993): the receptive skills of reading and listening rank higher than the productive skills of writing and speaking.

The reliability of the test is quite satisfactory: the coefficients are 0.94 (KR20) for the written test, 0.88 (KR20) for the listening test and 0.73 (KR21) for the oral test.

The level of difficulty and discrimination of the test is reasonable. The mean facility value of the written test is 0.640 with mean discrimination index of 0.314; and the mean facility value of the listening test is 0.612 with mean discrimination index of 0.526. All the sections of the tests are properly contributing to the total tests. The average item-test correlations of the sections range from 0.242 to 0.515 for both tests.

The validity of the test is also satisfactory. Evidence from the comments of students and language teachers suggests that the tests have good face and content validities.

There is evidence in support of the construct and concurrent validities of the tests. Both intercorrelations of sections and correlations between tests show a degree of common variance as well as some unshared variance; ie, with overlaps not exceeding 44% for the former and 62% for the latter. In other words, they are all testing English, but different aspects or skills are also being tapped as well. The greatest agreement ($r=0.81$) is observed between the new written test and the criterion (written) test. In addition, there is evidence of a clear relationship between the test scores and the University Semester Grade Point Average (SGPA): both the written test and the test of listening correlate with SGPA at $r=0.6$.

Finally, the study concludes with some observations on the testing of English at this University based on these findings.

ACKNOWLEDGEMENTS

I am greatly indebted to many individuals and institutions for the help and advice which made this study possible.

I am grateful to Dr. R. Hicks, Mr. J. Norrish and Dr. Dejene Leta for the comments and advice they gave me at the proposal and the early stages of my work. My sincere gratitude goes to Mr. D. Porter, University of Reading, for his critical comments and suggestions on the main work. He not only supervised my work closely but also gave me all the encouragement I needed.

Thanks are due to those many colleagues for the professional, technical and administrative assistance they gave me.

My thanks also go to the students who participated in the project.

I would like to record my thanks to the institutions that provided me with the necessary time and funds: these are the Addis Ababa University - particularly my Department, the Institute and the Graduate School, and the British Council - particularly the offices in Addis Ababa and Reading.

I am especially indebted to my family for all their patience and support.

Finally, any faults of presentation or errors of fact or opinion are, of course, entirely my own responsibility.

CONTENTS

	<u>Page</u>
Abstract ...	i
Acknowledgement ...	iii
List of Tables ...	xii

CHAPTER ONE

THE PROBLEM AND THE SCOPE OF THE RESEARCH

1.1 Introduction...	1
1.1.1 The English language in Ethiopia ...	1
1.1.2 The evaluation of language proficiency ...	3
1.1.3 The service courses ...	5
1.1.4 The service course examinations ...	5
1.2 Aim of the research ...	7
1.3 The scope of the study ...	8
1.3.1 Importance ...	8
1.3.2 Limitations ...	9
1.3.3 Organisation of the thesis ...	10

CHAPTER TWO

TESTING ENGLISH AT ADDIS ABABA UNIVERSITY: THE STATUS QUO

2.1	Freshman English 101A Examination ...	13
2.1.1	Reading comprehension ...	13
2.1.2	Vocabulary ...	13
2.1.3	Structure ...	14
2.1.4	Composition ...	14
2.1.5	Summary ...	14
2.1.5.1	Distribution of marks...	15
2.1.5.2	Item context ...	15
2.2	Freshman English 101B Examination ...	16
2.2.1	Reading comprehension ...	17
2.2.2	Vocabulary ...	17
2.2.3	Structure ...	17
2.2.4	Composition ...	18
2.2.5	Summary ...	18
2.2.5.1	Distribution of marks ...	19
2.2.5.2	Item context ...	19
2.3	Concluding remarks ...	20
2.3.1	Comprehension rubrics ...	20
2.3.2	Contexts ...	21
2.3.3	Question types ...	22
2.3.4	General ...	24

CHAPTER THREE

A REVIEW OF PRINCIPLES AND APPROACHES IN LANGUAGE TESTING

3.1	Language teaching and testing ...	28
3.2	Approaches in language testing ...	35
3.2.1.	The nature of language proficiency ...	36
3.2.1.1	Definition ...	36
3.2.1.2	The need for a theory ...	37
3.2.1.3	Non-unitary competence hypotheses - (one)...	39
3.2.1.4	The unitary competence hypothesis...	40
3.2.1.5	Non-unitary competence hypotheses - (two)...	43
3.2.2	Purpose oriented language testing...	50
3.2.2.1	Communicative language testing ...	50
3.2.2.2	Testing English for academic purposes ...	51
3.3	Some basic considerations in testing ...	54
3.3.1	Test validity ...	57
3.3.1.1	Face and Content validity ...	58
3.3.1.2	Construct validity ...	59
3.3.1.3	Concurrent validity ...	60

	<u>Page</u>
3.3.1.4 Predictive validity ...	60
3.3.2 Test reliability ...	62
3.3.3 Test efficiency ...	63
3.4 General remarks ...	64

CHAPTER FOUR

METHODOLOGY

4.1 The subjects ...	67
4.2 Instruments of data collection ...	68
4.2.1 The tests ...	69
4.2.1.1 The local tests ...	69
4.2.1.2 The external tests ...	70
4.2.2 The questionnaires ...	72
4.3 The construction of the instruments ...	73
4.4 Examining and marking ...	75
4.5 The data collection process ...	76
4.6 Analysis of data ...	78
4.6.1 Initial analysis ...	79
4.6.2 Main analysis ...	79
4.6.2.1 Construct validity ...	79
4.6.2.2 Criterion-related validity ...	79

CHAPTER FIVE

THE DESIGN AND CONSTRUCTION OF THE TESTS

5.1	Review of needs analysis ...	80
5.2	The specification ...	86
5.3	Realization of the specification ...	95
5.3.1	Description of the tests ...	97
5.3.2	Description of the subtests ...	102
5.4	Initial analysis ...	111
5.4.1	Descriptive statistics and test reliability ...	113
5.4.2	Item analysis of the tests ...	114
5.4.3	Test validity ...	118
5.4.3.1	Face and Content validity ...	118
5.4.3.2	Construct validity ...	123
5.4.3.3	Concurrent validity ...	124
5.4.3.4	Predictive validity ...	126

CHAPTER SIX

DISCUSSION OF THE RESULTS OF THE MAIN ANALYSIS

6.1	The revised tests ...	128
6.2	Descriptive statistics ...	129
6.3	Reliability ...	130

	<u>Page</u>
6.3.1 Test reliability ...	130
6.3.2 Marker reliability ...	131
6.4 The effect of background variables ...	135
6.5 Pass/Fail cut-offs ...	139
6.6 Item analysis of the tests ...	143
6.6.1 Classical item analysis ...	143
6.6.2 Rasch analysis ...	152
6.7 Test validity ...	170
6.7.1 Face validity ...	170
6.7.2 Content validity ...	176
6.7.3 Construct validity ...	179
6.7.4 Concurrent validity ...	185
6.7.5 Predictive validity ...	186

CHAPTER SEVEN

CONCLUSIONS AND SUGGESTIONS

7.1 Conclusions ...	188
7.2 Recommendations ...	194

BIBLIOGRAPHY ...	198
------------------	-----

APPENDICES ...	207
----------------	-----

	<u>Page</u>
Appendix 1: Question and value distributions of Freshman English 101A and 101B (1985/86 - 1990/91) ...	209
Appendix 2: Summary of frequency of language skills ...	211
Appendix 3: Freshman English Examination (Written Test) ...	214
Appendix 4: Answer key and marking guidelines (Written Test) ...	231
Appendix 5: Freshman English Examination (Listening Test) ...	236
Appendix 6: Listening text for listening test ...	240
Appendix 7: Complete text for partial dictation ...	242
Appendix 8: Answer key and marking guidelines (Listening Test) ...	243
Appendix 9: Freshman English Examination (Oral Test set) ...	244
Appendix 10: Examiners' set for the Oral Test: source and scoring sheet, all with instructions and evaluation guidelines ...	254
Appendix 11: External written test (WT2) ...	261
Appendix 12: Answer sheet for external written test ...	284
Appendix 13: Answer key and marking guidelines for external written test ...	286
Appendix 14: Answer sheet for external listening test ...	287
Appendix 15: Answer key and marking guidelines for external listening test ...	292

	<u>Page</u>
Appendix 16: Student Questionnaire for the Written Test (SQ1WT) ...	293
Appendix 17: Student Questionnaire for the Listening Test (SQ2L) ...	296
Appendix 18: Student Questionnaire for the Oral Test (SQ3S) ...	299
Appendix 19: Language Teachers Questionnaire for the three tests ...	302
Appendix 20: Item results for external written test (WT2) ...	310
Appendix 21: Item results for external listening test (LT2) ...	314
Appendix 22: External listening test (Test 16), cassette (LT2) ...	316

LIST OF TABLES

	<u>Page</u>
Table 1: Distribution of marks for 101A ...	15
Table 2: Distribution of marks for 101B ...	19
Table 3: Most frequent language activities ...	81
Table 4: Representation of skills in terms of language activities and macro-skills ...	91
Table 5: Final distribution of language skills for the macro-skills ...	92
Table 6: Table of specification ...	100
Table 7: Descriptive statistics and Reliability coefficients ...	113
Table 8: Students' response in frequencies and percentages ...	119
Table 9: Teachers' response in frequencies and percentages ...	121
Table 10: Correlation coefficients between tests ...	123
Table 11: Correlation coefficients between the new tests and the criterion tests and grades ...	124
Table 12: Correlation coefficients between the new tests and the criterion grade point averages ..	126
Table 13: Descriptive statistics and Reliability coefficients for the revised tests (KR20) ...	129
Table 14: Intra-marker correlation coefficients ...	131
Table 15: Inter-marker correlation coefficients ...	132
Table 16: Descriptive statistics by School, Town, Faculty, and Sex ...	135

Table 17: Computed F values from analysis of variance ...	137
Table 18: Results of item analysis for test components ..	144
Table 19: Results of item analysis for each item and section 146 ...	146
Table 20: Number of person and item misfits in the tests ...	155
Table 21: Rasch analysis of the written test (WT1) ...	159
Table 22: Rasch analysis of the listening test (LT1) ...	167
Table 23: Summary of views of candidates and teachers ..	171
Table 24: Summary of students' opinion on the components of the tests ...	175
Table 25: Summary of teachers' views ...	177
Table 26: Intercorrelation coefficients between tests ..	179
Table 27: Intercorrelation coefficients between the sections of the new tests ...	183
Table 28: Intercorrelation coefficients between tests: corrected for attenuation ...	184
Table 29: Intercorrelation coefficients between tests and grades ...	185
Table 30: Intercorrelation coefficients between the new tests and grade point averages ...	186
Table 31: Intercorrelation coefficients between sections and whole tests ...	187

CHAPTER ONE

THE PROBLEM AND THE SCOPE OF THE RESEARCH

1.1. Introduction

1.1.1. The English Language in Ethiopia

The English language has assumed a significant role as a foreign/second language in Ethiopia for at least half a century. The need for a foreign/second language arose from the introduction of modern education and the country's desire to establish contacts with the outside world (Dejene, 1990:11). Hence, two rival languages, ie, French and English posed to serve these purposes. Eshetie's assessment of the status of these languages in Ethiopia before the Italo-Ethiopian war is that "...the superiority of French as the language of instruction and diplomacy in Ethiopia, although challenged by English, remained unaltered until the Fascist occupation of 1935" (1974: 84). Dejene (1990: 12) also establishes the fact that "...English succeeded in assuming the role of Ethiopia's language of instruction and international contacts only after 1935, a period which marked the beginning of modern education in the country." Thus given this historical link between the English language and the introduction of modern education in Ethiopia, he goes on further to sum up that "... English has played a significant role in the educational system of the country ever since" (ibid.).

More specifically, the English language is given as a subject beginning from grade three at the elementary level, and serves as a medium of instruction from grade seven at the junior high school level (Ministry of Education, 1982:3).

Its use as a medium of instruction from junior secondary right up to tertiary levels of education, its role as a second official language of the country, its service as the official language of international organisations and some national organisations involved in international contacts and/or transactions illustrate the importance the English language has gained in this country.

Dejene (1990: 13) mentions attempts made to replace English by Amharic as a medium of instruction at the junior and senior secondary levels, and eventually at the tertiary level of education. But, he also points out the problems encountered; that is, difficulties in the translation of (subject) textbooks into Amharic in terms of the time and finance it would require to effect the undertaking. Thus he concludes:

Even if and when the change from English to Amharic is effected, it seems reasonable to believe that English will still continue to be taught as a subject since it will remain an important tool for Ethiopia's international contacts as well as a vehicle for importing science and technology. It is, thus, reasonably certain that the importance of English will continue into the future of Ethiopia (pp. 13-14).

Given the noticeable part the English language plays in the Ethiopian context, especially in the educational system of the country, the concern to study and explore better ways of teaching and/or learning the language as well as the evaluation of the outcome of this clearly becomes relevant.

1.1.2. The evaluation of language proficiency

Assessment in language is a necessary twin to language teaching. Knowledge of the extent to which an individual has learnt a language and can exploit it for communicative purposes is increasingly in demand. In the words of Bachman and Clark (in Bachman et al., 1988: 192-130).

There is an increasing need for measures of language proficiency... for making decisions regarding whether individuals have attained levels of language proficiency required for various educational and employment goals.

Thus, language tests for the assessment and evaluation of an individual for a given purpose.

Given the need for assessment and evaluation in educational contexts, language tests or examinations are administered at different stages of learning, with varying levels of complexity. Such language tests are often expected to indicate the candidates' command of English which would be useful in the academic endeavour they undertake. This is often true of tests taken before entry to or after the completion of language improvement course(s) in institutions of higher education where the English language is the medium of

instruction. In other words, language tests need to provide adequate information on the individuals' communicative ability in the English language which would help them to cope with the academic demands they are to face; and this directs attention to proficiency tests of English in general, and to tests of English for academic purposes in particular.

Where the English language is the medium of instruction, inadequate communicative ability in this language may have an adverse effect on the individuals' academic achievement in other courses.

Conversely, if individuals have a good command of the language, they may be at an advantage to do well in their academic performance. In the light of this, Criper and Davies (1988), in the ELTS validation research, found that language plays a role though not a dominant role, and Weir (1983), in the study of language problems of overseas students, argued that language is an important one though not the sole factor in academic success.

Given that language is important to pursue further and higher education, terminal English language tests should provide a profile of an individual in terms of what he/she is able or not, and contribute to the prediction of the likelihood of his academic success.

1.1.3. The service courses

Focusing on the local circumstances, the Department of Foreign Languages and Literature of Addis Ababa University offers English Language service courses to all students during their first and second years of university education. The overall aim claimed of the service courses is to promote the students' command in the use of English for academic endeavour.

The Freshman English is offered in two parts, i.e., 101A and 101B, during the first and second semester of first year. Both courses consist of contextualized exercises in reading comprehension, vocabulary, grammar and writing at the sentence and paragraph levels. The Sophomore English is a one semester course offered during either the first or second semester of the second year. It focuses on writing, i.e., paragraphs and essays, especially on the conventions of formal and academic writing. To achieve this end the exercises and activities involve reading comprehension, note-making, vocabulary, discussion, etc.

1.1.4. The service course examinations

The examinations of these service courses follow the nature and pattern of the courses. For the Freshman English courses the students take mid-semester tests and final examinations.

A typical Freshman English examination is composed of items like multiple-choice, fill in the blank, matching, etc. as well as writing a paragraph and rewriting some sentences. The main components of the examination are reading comprehension, vocabulary, grammar and paragraph writing. At the end of each course the students are awarded letter grades according to cut-off points decided on based on the normal curve.

In the light of the general dissatisfaction with the present quality of education, some doubt is cast on the effectiveness of the courses. Although there is no documented evidence, some people believe that the general decline in the quality of education is attributed, directly or indirectly, to the unsatisfactory performance of the students in English. Consequently, language achievement in the form of grades is distrusted.

With regard to language achievement, i.e., in terms of letter grades, there is a common expression of doubt whether they reflect any real language ability evidenced in academics. It appears that hardly any significance or meaning can be attached to these achievements. Given the nature of the language examinations and the way the letter grades are arrived at, the expressed achievements do not inform us about what the individual can and cannot do with language. Still worse, they are not even good indicators of some general and lasting standard or language capacity because the awarding of the letter grades is based on the performance of a population on a test at a given time.

In other words, the fact that the cut-off points for the letter grades vary each time adds to the doubt.

As regards the language examinations, very little or hardly any effort is made to check and control the quantity and the quality of either the individual items or the tests as a whole according to any set procedure. The only way these are checked is by the thorough discussion of the items or the sections, and so of the whole tests. Issues related to the examinations, i.e., the individual items, components or formats are raised, discussed, and changes are made where necessary to the satisfaction of the members of the testing committee. For example, the examinations are not pre-tested; the items are not analyzed for level of difficulty and discrimination index. In other words, the examinations have not been subjected to any set procedure to assess whether they measure what they are intended to measure and how well or how consistently they do this; i.e., there is no information on the validity and reliability of the tests. Furthermore, there is no information on the equivalence of the different forms of the test administered each year.

1.2. Aim of the research

Given the circumstances, it appears that there is need for improvement in the language testing scheme. Both the tests and the testing practice need to be updated in order to meet the demands of the situation. In general terms, the aim is to develop an appropriate and suitable test and testing

procedure for the circumstances outlined above: that is, testing English for tertiary education. In other words, the study attempts to identify the language needs of the Addis Ababa University students in order to develop a suitable test, thereby ascertaining whether a different approach to the testing of English at Addis Ababa University is advisable.

The intention is to redress the problems identified in the 1985/86 - 1990/91 examinations and develop a more purpose oriented test to the extent that it is appropriate to the context in which the English language is used.

The test, in its procedure and content, takes into account the language needs of the learners and reports performance in terms of language abilities or study skills. Thus, the information from such a test could be more directly meaningful to all parties involved.

1.3. The scope of the study

1.3.1. Importance

Such a test is hoped to have a desirable impact on the language programme, i.e., regarding its basic approach to language teaching and/or the organisation (structure) of the service courses in the university. Differently put, the backwash effect may result in some positive changes that would increase the effectiveness of the courses.

Such a study, especially the findings of the study, may also suggest the desirability of a separate university entrance examination, instead of relying on the Ethiopian School Leaving Certificate Examination (ESLCE); at least, the results may provoke questions and hint at the need for further study in this matter. Furthermore, the study might show, on pedagogic and economic grounds, the desirability of a placement test.

1.3.2. Limitations

At the dawn of the opportunity, this topic was conceived based on the belief that a research gap existed in the then new Ph.D. research project concerns. In our department, a person was working on a project related to the assessment of English language at high school completion level (ESLCE), and two others were working on projects related to English language syllabus design at university level. Thus, evaluation of English for tertiary education was considered a research gap that needed to be filled in. This project beyond high school completion and as a necessary adjunct to the syllabus design, was proposed to satisfy this need. However, this researcher, given the sometimes biting demands caused by the extent of the topic, now feels that it would have been appropriate to narrow it down.

The specification was drawn up from a needs analysis conducted in 1982 and another in 1991/92. The former is doubted both for its age and purpose whereas the latter only for its

purpose (See section 5.1). Had it not been for fear of duplication and financial constraints, a fresh needs analysis would have been proper and appropriate.

In the initial proposal, it was planned to involve about 1200 students (subjects). But, due to financial limitations again, the present analysis is based on the reactions of about 256 (Gross) students and 19 English language teachers. Clearly, a larger sample size than this would have been preferable, thereby allowing for more confident generalizations to be made about the target population.

1.3.3. Organization of the thesis

In this chapter, we looked at the status of the English language in Ethiopia, the need for the assessment of language proficiency, and the scope of the study in terms of its significance and the limitations within which the research is conducted. The aim of the research together with the problem it attempts to solve has been dealt with.

Chapter Two is devoted to a detailed description of the local problem and/or the practice of testing English at Addis Ababa University. The nature and format of six years Freshman English examinations is described.

In Chapter Three a brief review of the relevant principles and approaches in language testing is presented. The developments in language testing are traced and research undertakings of

similar concern are reviewed. Some basic considerations that need to be taken into account when designing a suitable test for a given purpose are mentioned.

A brief presentation of the methodology adopted in this work is given in Chapter Four. The instruments used, the methods or procedures followed in their construction as well as in the collection and analysis of the data is provided.

Chapter Five deals with the detailed procedures followed in the design and construction of the tests. It includes a review of needs analysis, the specification and its realization in terms of the contents and methods of the tests. This description of the tests is followed by the analysis of the initial administration of them.

The discussion of the results of the main analysis of the final administration of the tests is presented in Chapter Six. Finally, conclusions and suggestions are offered in the last chapter.

CHAPTER TWO

TESTING ENGLISH AT ADDIS ABABA UNIVERSITY: THE STATUS QUO

An exploration into six years (1985/86 - 1990/91) of Freshman English (for each of the courses FLEn 101A and FLEn 101B) final examinations for regular (full-time) students has been carried out to provide an overview of the examinations (Appendix 1). This has been done, for instance, in terms of the distribution pattern for the four components, (ie, comprehension, vocabulary, structure and composition) for objective/subjective categories of question types, and the contexts in which test items were presented. In other words, the Freshman English examinations are described and critical comments are also offered.

While the questions that required the candidates to choose from a given set of alternatives and the scoring of the response was made with reference to an answer key were classified under the objective category, open-ended items, the marking of which depended on the personal judgement of the marker, were classified under the subjective category of question type. Items of true or false, matching, labelling, multiple-choice and its variants, etc. were considered objective whereas completing sentences, writing (re-writing) full sentences and paragraphs, etc. were considered subjective question types.

2.1. Freshman English 101A Examination

Considering the total number of questions and the marks of the final examinations, Freshman English 101A had 94 questions for the first year, 93 for each of the subsequent three years, then 86 and 75 questions for the final two years, for the value of 60 marks for each of the first four years, then 55 and 50 marks for the last two years respectively.

2.1.1. Reading comprehension

While comprehension constituted 25 per cent of the total marks for the first four years, it rose to 27 and 30 per cent in 1989/90 and 1990/91 respectively. We also notice the number of passages used became two since 1988/89. The proportion of objective/subjective remained more or less even, except between 1986/87 to 1988/89 when the subjective rose to 17 per cent and objective dropped to eight per cent. In 1985/86 and 1989/90 it was one and two per cent, respectively, in favour of objective type, and in 1990/91 it was two per cent in favour of subjective type.

2.1.2. Vocabulary

This component constituted 33 per cent of the total marks for the first four years, but dropped to 18 and 20 per cent in the last two years, respectively. Here, we also notice that the subjective section "Construct meaningful sentences with the given words..." was dropped after 1989/90. With respect to

the ratio of objective/subjective types of questions the component consisted of 25 per cent objective and only eight per cent subjective for the first four years, but it all became objective only for the last two years.

2.1.3. Structure

This took up 33 per cent of the total marks for the first four years, went up to 36 in 1989/90, and then dropped to 30 per cent in 1990/91. It is observable that the objective type fluctuated in the range of 24 to 31 per cent of marks whereas the subjective did the same in five to eight per cent range. In the former, it was 25 per cent in 1985/86, 28 for the next three years, then it rose to 31 in 1989/90 to decline to 24 per cent for the subsequent year. And in the latter, it was eight in 1985/86, then five for three years before it rose to six per cent in the final two years.

2.1.4. Composition

This was eight per cent of the total marks for the first four years, but rose to 20 per cent for the last two years; that is, it rose from one writing task to two tasks.

2.1.5. Summary

Overall, the structure component in Freshman English 101A received the greatest emphasis in terms of mark allocation,

followed by comprehension and vocabulary very closely. But gradually comprehension ascends to a clearly second position.

2.1.5.1. Distribution of marks

Table 1: Distribution of marks for 101A

Year	Objective (%)	Subjective (%)	Total
1985/86	63.3	36.7	100
1986/87	61.7	38.3	100
1987/88	61.7	38.3	100
1988/89	61.7	38.3	100
1989/90	72.7	27.7	100
1990/91	68	32	100

Freshman English 101A is a mixture of objective and subjective type of questions, with a predominance of the former. It was 63 per cent objective in 1985/86, and 62 per cent in favour of objective in the following three years. In 1989/90 this proportion rose to 73 per cent, and fell back to 68 per cent in the final year. The table also suggests that the proportion of the objective type is on the rise.

2.1.5.2. Item context

Considering context in the vocabulary and structure components, the examinations predominantly involve performance at the sentence level. Focusing on the former, we see that the first four years of the examinations present the exercises in sentence contexts with the exception of prepositions,

which are supplied in a paragraph context. That is, of the 33 per cent devoted to vocabulary, 75 per cent is in sentence contexts. During the latter two years, however, the task changes to one of matching synonyms and antonyms for words taken from the passage.

A different picture emerges for the structure portion, that is, the predominance of sentence contexts seems to have terminated in 1985/86. Of the 30 to 36 per cent devoted to structure, there was 38 per cent paragraph context in 1985/86, but in the next three years this rose to 60 per cent, and in the last two years it was all in paragraph contexts with one short section in more than one sentence context.

In the composition, candidates were required to write only one short paragraph for the first four years, but they had to do two writing tasks in the last two years, that is, besides the paragraph, they had to write complete sentences from given sentence fragments. Regarding the paragraph writing, candidates were provided with one sentence (sometimes incomplete) cue, or with notes.

2.2. Freshman English 101B Examination

Regarding the total questions and the marks of the examinations, Freshman English 101B had 69 questions for the first two years, 71, 62, 72 and 83 questions in the following consecutive years, for the value of 50 marks for the first

four years, 45, and 55 marks in the last two consecutive years.

2.2.1. Reading comprehension

This constituted 30 to 36 per cent of the total marks of the examinations. We also observe that for the latter two years the number of passages used rose to two instead of one. In terms of the objective/subjective proportion, we notice some fluctuation (with a slight decline in the middle) in the subjective type, and a clear rise in the objective through the six years. In the first two years subjective questions were 24 per cent, in 1987/88 and 1989/90 20 per cent, in 1988/89 22 per cent, and in 1990/91 16 per cent. On the other hand, the objective questions were six per cent for the first two years, ten in 1987/88, 11 in 1988/89, 14 in 1989/90, and 20 per cent in the last year.

2.2.2. Vocabulary

This took up 20 to 22 per cent of the total marks of the examinations: that is, it was 20 for four years, and 22 per cent in 1988/89 and in 1990/91. However, it was only objective type for all the six years.

2.2.3. Structure

The total value for this component shows a progressive decline, that is, from 40 per cent for the first three years

to 33, 26, and 23 per cent in the latter three consecutive years. The ratio of objective/subjective was evenly distributed at 20 per cent each for the first three years, but began to fluctuate during the next two years with alternating emphasis on each type. In 1988/89 there was a difference of seven per cent in favour of subjective type whereas there was a difference of 14 per cent in favour of objective type in 1989/90. And it was only objective type in the final year.

2.2.4. Composition

This component was ten per cent of the total marks for the first four years, although it went up to 20 per cent during the last two years, that is, it rose from one to two writing tasks.

2.2.5. Summary

overall, the structure component in Freshman English 101B received the greatest emphasis, followed by comprehension for the first three years. In 1988/89 both components received equal emphasis (attention) at about 33 per cent of marks. In the last two years, however, the situation has been reversed - that is, the greatest emphasis has been given to comprehension followed by the structure.

2.2.5.1. Distribution of marks

Table 2: Distribution of marks for 101B

Year	objective	Subjective	Total
1985/86	46	54	100
1986/87	46	54	100
1987/88	50	50	100
1988/89	46.7	53.3	100
1989/90	54	46	100
1990/91	65.4	34.6	100

Freshman English 101B is also a mixture of objective and subjective types of questions, but the proportion of these is more or less evenly distributed with the exception of the last year. While for the first two years the distribution is in favour of subjective by eight per cent, it levels off to an even distribution in 1987/88, rising again to seven per cent in the following year. However, the proportion is reversed for the last two consecutive years - that is, it is eight and 31 per cent in favour of the objective type. This table reveals that the predominance of the subjective type shifts to the objective, and that this latter one is definitely on the rise, perhaps with great leaps.

2.2.5.2. Item context

Concerning context in the vocabulary and structure components, again the examinations predominantly involve performance at the sentence level. This is particularly so in the first four years with the exception of the use of a paragraph context to

supply prepositions in 1987/88 and 1988/89, which accounts for 13 per cent of the total. The predominance of paragraph context clearly emerges during the last two years in both components. Especially in the vocabulary, candidates were required to match meanings with words they had encountered in the passage instead of choosing the appropriate word to fit a given sentence as previously.

In the composition, candidates were required to write one paragraph from given notes for the first four years, but in the last two years they had to do two writing tasks, that is, writing a paragraph from notes and another from a given sentence cue.

2.3. Concluding remarks

2.3.1. Comprehension rubrics

A survey of the rubrics for the subjective questions in comprehension in both Freshman English 101A and 101B reveals that the purpose of the component is to evaluate the candidates' ability to understand, and express that in accurate sentences; that is, the task is loaded with a complex purpose. And this could be a cause for some disparity in marking. For example, those markers who see only one purpose (i.e., comprehension) in the exercise awarded full value for a correct answer in poor English whereas those who see a double purpose (i.e., comprehension and accuracy) awarded half the value for the same response. In a speculated extreme,

those who see the structural accuracy of the response as the purpose might award half or even full value to a response impressed by the language or the language potential of the candidate. On the contrary, adherents of the other extreme might not reward the same response at all. Typical phrases in the instruction include: "Answer... in your own words in single sentences,... in complete and correct sentences,... fully in complete and correct sentences,... in one or two sentences,... according to information in the passage."

Furthermore, this purpose (i.e., assessing accuracy of sentences), and this problem (i.e., the disparity in marking) run through, to a lesser or greater degree, the vocabulary, the structure, and composition components. Examples of these are: the "Construct grammatical and meaningful sentences using the following words..." in the vocabulary, the "Put/Change the following into reported speech..." in the structure, and the "Put the following fragments in correct order to make meaningful sentences..." in the composition.

2.3.2. Contexts

While there is an observable (and desirable) move towards larger contexts like paragraphs, the examinations involve predominantly sentence level contexts which offer less authenticity and meaningfulness than paragraphs. This is, of course, a reflection of the course (the teaching): that is, the fact that language items (elements) are dealt with discretely or in isolation.

The move towards paragraph or passage contexts, for example in testing structure, is also questionable; that is, one wonders whether length should not be thought about. Some contexts certainly appear to be rather longish, and whether they can maintain interest (motivation) to the end is doubtful. Length in itself, among other things, might bar the candidate from showing his full capacity. Providing more than one shorter but complete context has the advantage of offering a variety of topics which would increase the candidate's opportunity to find something that would interest him/her rather than in just one long passage. He/she would also have the pleasure of beginning afresh which would satisfy his/her sense of progress through the test instead of feeling stuck or lost in one long passage.

2.3.3. Question types

Both Freshman English 101A and 101B happen to be a mixture of objective and subjective type of questions, but mainly in favour of the former, overall. While it may be difficult to argue about a consistent and deliberate pattern, especially in 101B, in this distribution over the six years, it is probably safe to say that the proportion is rather less erratic. This may be so partly due to the desire, as expressed on some of the Testing Committee meetings, to maintain some sort of a standard (at least in form) in the examinations and to achieve some degree of uniformity in grading.

Also, whether there are enough open-ended questions which require students to produce language remains questionable: especially when this is viewed in the light of the level of difficulty a cloud of doubt remains in the air. This is, of course, assuming that open-ended questions are more authentic and demanding to respond to than the receptive ones. In other words, whether the two courses are graded (at least in this respect), with Freshman English 101B at a higher level, or whether the two examinations are testing ability at different levels is not very clear. This, of course, needs to be viewed against the fact that success in the first course is required to register for the second course.

Objective (receptive) questions, e.g. multiple-choice items, are tests of recognition, and even if we accept that they may test this ability at varying levels of difficulty, they remain limited to the testing of this ability. Given that about 50 or 60 per cent of the examinations are objective questions, it is difficult to be certain that these examinations test a variety of relevant abilities or skills. And whether there is a steady relationship between the productive - receptive skills is not very clear.

And besides, these objective questions mainly encourage rote memorization, especially so in the vocabulary and structure components. For example, in the vocabulary candidates are presented with a list of words (from the textbook) to choose from and use in the one sentence contexts. Clearly, the responses to these call for memorization as a study habit.

During the latter two years, however, the tasks change to one of matching synonyms and antonyms for words taken from the passages in the examinations. This exercise tests the ability of the candidate to work out the meanings of words from the contexts in which they are used. And this is a realistic reading skill and a test of a reading strategy. Also, in the structure, candidates are required to change the forms of words (verbs) or choose one from the alternatives given in brackets in a larger context than the sentence. While the use of a larger context is desirable, the purpose remains a test of accuracy; more specifically, a test of accuracy of form in a meaningful context. This, too, promotes memorization as a study habit.

This habit is further advanced also in the composition: the one or incomplete sentence cues provided for paragraph writing are based or related to topics in passages which are dealt with in the textbooks or in class. Even the paragraphs written from notes too, appear to be rather awkward during marking for most of them are made up of juxtaposed notes. They do not read smoothly for they lack some internal linking. Besides, the writing task does not seem to be related to academic writing like making notes, summarizing, etc.

2.3.4. General

We have noticed changes in the number and type of questions (items) and the context in which they were presented. These changes, often motivated by individuals in the Testing

Committee, occurred in spite of the occasional call to maintain some standard and uniformity in all aspects of the testing scheme. Although one can see the tacit consensus about some of the changes, it is rather difficult to say whether the changes have been adopted (formally) as a departmental policy. If there was such a policy, a more permanent pattern (picture) would have been observed.

The test items (exercise types) are hardly related to academic activities be it in content or in form. They hardly appear to be testing any study skills either. The fact that structure and usage generally receive the greatest emphasis is clear evidence that the orientation of both the courses and the examinations is structural. Hence, the examinations are predominantly achievement-oriented and structure-based. Moreover, this is also evident in the course descriptions of the department (1988: 24), and in the stated aim of the textbooks (1985: i). The course descriptions read as follows:

FLEn 101 (101A in the textbook) Freshman English (3 credits)

A revision course in which students practice the fundamentals of English Structure, (i.e., present tense, past tense, future tense, reported speech, nouns and articles). It includes reading comprehension and vocabulary work.

FLEn 102 (101B in the textbook) Freshman English(3 Credits)

Comprehension: vocabulary building (lexical work, word building, affixes and noun compounds): idiomatic usage, phrasal verbs, and prepositional phrases; structure and usage (i.e. relative clauses, conditional sentences, infinitive and gerund, passive voice, and sentence combination): reading; simple paragraph writing.

The stated aim in the textbooks is slightly differently phrased, and it reads:

... The aim is to review and revise the ground rules of English structure and usage while at the same time extending the students' active and passive command of the language. Thus, it is intended to be an aid to improving the students' academic efficiency in their university studies....

This statement doesn't appear to be committed to one particular approach. Instead, we notice that the books are primarily based on the structural approach but with an expressed intention of English for academic purposes. Clearly, there is a mismatch between the means and the intended end, and thus, there is little hope that the means would lead to the desired end. This is a prime issue that needs to be redressed. In fact, this mismatch between the stated objectives and the means of achieving them indicates some ambiguity which provides a reason to suspect the effectiveness of the service programme. In other words, the courses seem to lack a clear definition of objectives and the means with which to achieve them. And this state of affairs is carried over to the assessment scheme. For instance, the appropriacy and effectiveness of the language programme, and hence its assessment, for science students is often questioned. There have been complaints that many students had to be dismissed because they had failed, not in the science subjects, but in one of the service courses. Note the severity of the situation: students enter academic warning or probation, or still worse, face academic dismissal because

they have done poorly in a course whose aim is to help them succeed in their other academic subjects.

Broadly speaking, the aim of the examinations, including the purposes of each component and most of the items, boils down to the assessment of general English, and definitely structurally oriented at that. In this connection, the assessment of structural accuracy takes the lion's share of the total value for each examination.

CHAPTER THREE

A REVIEW OF PRINCIPLES AND APPROACHES IN LANGUAGE TESTING

In this chapter, a brief review of the more recent principles and approaches as well as some basic considerations in language testing is presented.

3.1. Language teaching and testing

The general relationship between language teaching and testing is better put by Underhill (1982: 18) when he says, "AS YE TEACH, SO SHALL YE TEST". This statement directly refers to the different kinds of tests developed and administered variously for different purposes; that is, tests may be given for placement and remedial purposes as well as to assess the progress made and the extent of success achieved in the learning of the language. So, while the reference to achievement tests is rather direct, it is less so to proficiency tests.

But in this study, the development of a suitable proficiency test is sought in the context of Addis Ababa University. First, the ultimate goal of any formal instruction in the target language is to approach a certain level of proficiency, and the testing that goes with it is to measure the level of proficiency achieved. Ethiopian students study the English language as a subject from third grade at the elementary level up to the first year of their tertiary education. They also

study the other subjects in the English language from seventh grade at the junior secondary level right up to higher institutions of learning (Ministry of Education, 1982:3). So it appears that the test or examination that they take at the end of the last English language course should measure the proficiency level they have strived to achieve all along; that is, the proficiency level gained as a cumulative effect of a sequence of instruction in the language needs to be assessed rather than what has been achieved during the final English courses offered at the Addis Ababa University.

Second, recall the intention to teach English for academic purposes (ie. in the absence of an EAP course) as the stated aim in the textbooks for the English courses at Addis Ababa University (See Chapter II). This too, calls for the use of proficiency tests, more specifically, tests of English for academic purposes. Proficiency tests measure general language ability or skill, could provide a profile of an individual in terms of what he/she is and is not able, and make a contribution to the prediction of the likelihood of his/her academic success. However, this task has been addressed, deliberately or otherwise, to achievement tests, which by definition are meant to measure what students are thought to have learnt, based on a structural approach to language teaching and testing at this university. There is hardly any literature that confirms, or at least suggests, the notion that structure based achievement tests could make a contribution to the prediction of future performance.

Third, the heterogeneous nature of the students at Addis Ababa University adds a point in favour of a proficiency test rather than for an achievement test. The students come from different language, cultural, social and academic backgrounds. During and at the end of their English language courses they hope to use it in different fields and disciplines. In this connection, discussing tests, for example TOEFL & ELTS, intended to screen candidates wishing to pursue higher education in English speaking countries, Alderson (1988: 17) writes:

These heterogeneous future situations require a proficiency test rather than an achievement test. There may be cases where achievement tests can be used for predicting future behaviour but that would only be when the students are from a common background, and their future is common - there is no difference in their future target situations. Normally this does not apply and therefore one needs proficiency tests because of heterogeneous background or heterogeneous future.

Fourth, one final reason for wanting to experiment on a proficiency test within the circumstances prevailing at Addis Ababa University is related to the desired impact of the test - the backwash effect. The following quotations give rise to the thought. According to Alderson (1988: 17), "... what proficiency testers do is what syllabus designers do. Proficiency test writers have to design their own syllabus, since a proficiency test is not based on a particular syllabus." Hughes establishes the link between the final achievement test and a proficiency test based on his experience when working on a new testing system at Bogazici University. He (1988a: 39) argues:

In brief, they (course objectives) are to raise the standard of the students' English in such a way and to such a level that they will be able to meet the language demands made by their undergraduate courses, as established by the needs analysis. This is what it is hoped that the students will achieve; and the final achievement test is intended to discover how successful they have been. Since the proficiency test itself is based directly on the needs analysis, it is only to be expected that the two tests should have the same content and structure.

What is suggested is that the objectives of courses, the syllabus (as a plan for achieving objectives), and proficiency tests need to be established by needs analysis; and an achievement test based on such course (objectives) would have the same content as a proficiency test. So, in the absence of course objectives drawn up from needs analysis as at Addis Ababa University, a proficiency test would be preferred. If, furthermore, the aims and objectives of a proficiency test are drawn up from needs analysis, as is the case in this study, a desirable backwash effect can be anticipated, or at least hoped for. Broadly speaking, the proficiency test can serve as a sample of the proper course objectives, contents, structure, and sample of the language demands candidates will face - a miniature sample of these and the procedures involved for any subsequent endeavour to improve the English language courses.

This last point takes us into the more recent concern and/or belief in language testing - that is, beyond the concern to develop appropriate and accurate measures of language and beyond the belief that "The good test is an obedient servant

since it follows and apes the teaching" (Davies, 1968: 7), to the innovative stance of developments' in language testing. Nowadays, the concern is both with successful teaching and successful testing as expressed by Hughes (1988 a: 42): "The proper relationship of testing and teaching is surely one of partnership. And one function of testing is to provide the kind of information that will help keep its partner on the right track."

Examples of test developments, as instances of the spirit or belief expressed above, could be cited; that is, tests that have been informative and innovatory as to cause desirable changes and amendments in the teaching/learning situation. The common concern in most of these undertakings is to develop an appropriate test or testing system that would hopefully exert beneficial backwash effect.

Tests designed and developed by Johnson and Kin - Ling (1981) and Milanovic (1988) for two different institutions in the context of Hong Kong; the National Certificate in English in Sri Lanka (Alderson et al., 1987); those developed by Henning et al. (1981) at Cairo University and Hughes (1988) at Bogazici University; and the Test of English for Educational purposes (TEEP) by Weir (1983) are some that stand as good examples.

The desirable backwash effect of some of these efforts are already reported. For example: The new test designed by

Johnson and Kin - Ling and which was introduced into the secondary school system of Hong kong eventually:

Led to syllabus revision, the design of new textbooks with different goals and objectives from those currently in use, and as a consequence new class-room methods and techniques which stress the use of authentic English in a purposeful and interactive way (Pearson, 1988: 100).

However, whether these subsequent changes have actually improved the students' English is not as explicitly reported.

Also at Bogazici University the backwash effect, according to Hughes (1988b: 143-145), took the form of advisory visits of specialists on the selection of texts, methods, materials production and teacher training, drawing up a new syllabus, introducing new textbooks, and increased classroom hours and organizing a summer school for the weakest students. And the cumulative effect of all these was a marked improvement in the standard of English. In the list of lessons to be learnt from his experience in Turkey, Hughes (1988b: 145) concludes:

The second lesson concerns the test's considerable impact on teaching and learning It is my belief that this was due almost entirely to the fact that the test was criterion referenced and based directly on the English language needs of students. In these circumstances, teaching for the test (which may be regarded as inevitable) became teaching towards the proper objectives of the course.

In the evaluation report of the National Certificate in English in Sri Lanka, the project members (1987: 12) talk about information hand-out together with a model paper. This

document was produced to give full information to the public: it included the purpose of the new examination, the syllabus it intended to cover, detailed information about each component of the examination and suggestions as to how candidates should prepare for the examination as well as examples of the testing techniques. This mechanism, one would hope, may awaken or alert the public to the demands of (innovative) changes in the testing of English, which in turn is bound to influence their view of language learning, the way they actually learn the language and the way they prepare themselves for the examination. They also state that "The design and trialling of the new examination required specialised manpower, and it was necessary for the organisers of the NCE to establish an intensive staff development programme" (p.14). In 1984 and 1985 a total of 17 secondary school teachers were trained in the testing of language skills and the evaluation of test items and testing procedures.

The publication of a 'Teacher's Guide' is mentioned too:

This guide would consist of a detailed syllabus for the NCE, rationale and descriptions of all NCE components, and a collection of activities that could form the basis for self-study or teacher-directed learning. The authors ... intended that the publication ... would also encourage teachers to use activities in the classroom that would encourage useful learning and ... examination practice (p.14).

This too, in practice, is what Hughes (1988b) believes is 'teaching for the test' which, he considers, 'may be regarded as inevitable'. This situation also suggests that a proficiency test eventually becomes an achievement test.

Generally, the training of teachers, the publication of the information hand-out and the teacher's guide are meant to pave the way for a positive backwash effect, thereby carrying over the innovatory testing techniques into the teaching/learning situation.

However, a recent study (Wall and Alderson, 1993) on the washback effect of the Sri Lankan English examination reveals that only some modest success can be achieved in this regard. Although they have found evidence of washback effect on the content of teaching and on the way tests are designed, not all of this was found to be positive. In their own words:

Our conclusions are that the exam has had impact on the content of the teaching in that teachers are anxious to cover those parts of the textbook they feel are most likely to be tested ... There is no indication that the exam is affecting the methodology of the classroom or that teachers have yet understood or been able to implement the methodology of the textbooks (ibid., p.67).

These innovative tests, especially those by Milanovic, Hughes, Weir, and the Sri Lankan one, have been exemplary for this work and have inspired much of the thinking in this study.

3.2. Approaches in language testing

While the concern is with the nature of language proficiency in general, it is examined in the light of English for academic purposes in accordance with Cummins (1984) suggestion that language proficiency should be considered in its relation to the development of academic skills.

Bachman and Clark (quoted in Bachman et al., 1988) point out the inadequacies of current measures of language proficiency which are needed for various purposes.

In their own words (pp. 129-130):

Most currently available measures are inadequate... in two respects; they are based on a model of language proficiency that does not include the full range of abilities required for communicative language use, and they are based on norm-referenced principles of test development that only permit interpretation of ability levels relative to performance of specific groups of language users.

3.2.1. The nature of language proficiency

3.2.1.1. Definition

Language tests often dealt with the assessment of the knowledge of linguistic rules prior to the view that proficiency tests should be concerned with the language behaviour of the individual in a communication situation he might find himself/herself. In terms of the developments in language testing, Weir (1983: 25) traces the definition of language proficiency thus:

Traditionally it has been defined in terms of performance in tests of linguistic competence, which assess the ability to produce grammatical sentences or utterances through a knowledge of linguistic rules. The more balanced of recent developments in approaches to language testing regard proficiency as a matter of communicative as well as linguistic competence, because the effective control of English in an

appropriate situation requires command of use, as well as usage (contextual as well as linguistic or grammatical competence).

Proficiency tests tend to look forward (Alderson, 1988: 16-17; Heaton, 1988: 172-173) in that the candidate's ability is defined in terms of a future realistic situation, without reference to any syllabus. In other words, the interest is in measuring the candidate's command of English to perform in a given situation. Thus, language proficiency elicits the individual's language facilities useful in dealing with language demands he/she is likely to face in a particular context. The task he/she is required to perform may fall within the range of basic survival activities and those of an academic nature, i.e., a range from everyday interpersonal interactions to an appreciation or expression of serious thought and reflection in academics.

Do proficiency tests, then, satisfactorily measure what they are intended to measure? From the literature, it appears that many problems need to be solved before a definitely positive answer is offered to this question.

3.2.1.2. The need for a theory

The problems arise from the diversity of approaches and theories adopted, and the methods employed in test constructions as well as the appropriacy of statistical techniques used in offering empirical evidences in the validation of the measures. Bachman et al. (1988: 129)

acknowledge the amount of research carried out in the area of language proficiency. However, they comment that:

... while the result of this research and development has been an enriched pool of measures, the potential for genuine progress in the understanding of foreign/second language abilities has been limited by the diversities of approaches and theoretical frameworks that have been employed.

Skehan (1988: 211) too, observes:

Occasionally it is suggested that the problem of language testing could be solved by following one's students around for some considerable time, and observing how they perform in a very wide range of actual communicative situations, thus allowing powerful generalisations to be made about proficiency.

However, he with other scholars like Clark (1978), Bachman (1990), Shohamy and Reves (1985), Spolsky (1985) (cited in Skehan, 1988: 211), considers this proposal impractical, biased, unethical and unreliable. Although Skehan acknowledges the wide-ranging and realistic data collection, he questions the representativeness of the data on the grounds that length of time to pursue the student, and which situations to include are not known; thereby, demonstrating the central issue in language testing, i.e., the 'what and how' to sample. Moreover, he asserts that the decision on what and how to sample calls for a theory (ibid.):

... 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.

The need for theory is further confirmed by Vollmer (1983: p.29):

... it is actually more theory that is wanted and needed throughout. For only that which is being put into the construction of a particular test or set of tests by way of theoretical insight, reasoning, or hypothesis building in advance will come out of it substantially. It is a heavy burden and there is a long, long way to go [Author's emphasis].

While scholars like Chomsky (1989), Hymes (1972) and Oller (1979), for example, (cited in Canale, 1984: p.29) advocate the idea that language proficiency is comprised of underlying abilities, knowledge systems, and skills, there is generally less consensus among scholars with regards to the content and boundaries of this underlying competence and therefore on what is to be measured by language proficiency tests.

3.2.1.3. Non-unitary competence hypotheses - (one)

Continued research into the nature of language proficiency reveals that much effort has been and is being expended to capture a clear and accurate picture of the ability of a student in a foreign language. As a result, two major views or hypotheses have so far been forwarded: the unitary as against the non-unitary hypothesis models -- ranging from one global component to many separate components of language proficiency, with claims of theoretical plausibility and support from empirical evidence (Vollmer, 1983: 3; Cummins, 1984: 5). And each of these approaches are based on differing hypotheses about the nature and structure of linguistic and communicative competence.

Vollmer (1983: 6) observes that language testing has been based on a multi-dimensional approach since the work of Fries and Lado because of the influence of the linguistic theory of structuralism. Accordingly, knowledge of a language was understood as mastery of its elements, its codes and their manipulation for use. So, language ability was understood as having several distinct aspects which were believed to be identical with the dimensions of attainment in a foreign language that could correspond to a performance matrix (ibid.). He goes on to say that (ibid.):

From the point of view of classical test theory it was argued that the degree of mastery in a foreign language could be inferred from the multitude of measured language performances (at least four). Thus it would be possible to form some sort of overall picture of a person's language ability and ... of his ability to act in a more or less predictable manner even in future situations requiring language use

Regarding these classical multi-dimensional models, the same writer points out that their theoretical claims were not explicitly stated, and that they drew upon linguistic, psycholinguistic or psychometric theory depending on what was available during formulation (p. 7).

3.2.1.4. The unitary competence hypothesis

As opposed to the multi-dimensional model, there arose the Unitary Competence Hypothesis (UCH). In tracing the theoretical basis for the assumption of a unitary competence, Vollmer (1983: 9) detects that the shift of focus occurred in

the work of Spolsky (1973: 173-174) who defined knowledge of language as mastery of rules, that is, an 'underlying linguistic competence', which in turn manifests itself in all the various kinds of performances-active or passive. For Spolsky (1973: 174) knowledge of the codes of a language is the 'principal factor' in the perception and production of messages, although this is not the sole factor accounting for all kinds of language behaviour. So, tests of passive or active skills would provide useful information about a candidate's knowledge of language, although the individual's performance as a speaker cannot be equated to his performance as a listener. But the claim is clearly "... that the same linguistic competence, the same knowledge or rules, underlies both kinds of performances" (ibid.). From this statement Vollmer (1983:9) asserts that:

This ... clearly indicates the shift of focus that has taken place from the differences between the skills (and how they might relate to underlying competence) to what they might have in common by way of a shared basic competence stretching out into all the skills. It is not claimed (as yet) that all possible linguistic performances are based on one and only one single ability [Author's emphasis].

Oller (1979) and Oller and Perkins (1980), as an advance in the same direction, describe the Unitary Competence Hypothesis. Oller and Perkins (1980: 1) claim that:

a single factor of global language proficiency seems to account for the lion's share of variance in a wide variety of educational tests including nonverbal and verbal IQ measures, achievement batteries, and even personality inventories and affective measures ...

It was thus proposed by Oller(1979) (quoted in Skehan, 1988: 212):

... that there is one underlying competence for language behaviour, based on the learner's pragmatic expectancy grammar ... one can only comprehend (and produce) language by means of process of analysis-by-synthesis, i.e., we comprehend by predicting the message we are encountering, drawing upon our knowledge of the world and, through the pragmatic expectancy grammar, the redundancy built into language.

This grammar is most directly operationalized by integrative tests commonly characterized by cloze and dictation (ibid.: Heaton, 1988: 16). These are tests of language in context focusing on meaning and the overall communicative effect of discourse that aim at measuring the candidate's ability to employ more than one skill simultaneously (Heaton, 1988: 16). It is also further claimed that every learner possesses this grammar of expectancy regardless of the purpose for which language is being learnt (ibid.).

The formulation of the UCH is partly inspired by findings in cognitive psychology where the role of the active and constructive perceiver is described as 'analysis-by-synthesis' (Vollmer 1983: 10).

This approach, however, was criticized much on the basis of its theoretical ambiguity (Vollmer, 1983: 12) and the inappropriate use of the factor analysis technique (Vollmer, 1983: 14; Skehan, 1988: 212), in establishing empirical evidence. Vollmer (1983: 17) quotes Oller where he admits"

that the strongest form of the unitary hypothesis was wrong" [Oller's emphasis].

The views or issues explored so far (in this section) generally relate to the psychometric-structuralist and psycholinguistic-sociolinguistic eras (Spolsky, 1975; Weir, 1983) in the recent history of language testing; and these are characterized as 'the Vale of Tears' by Morrow (1981). While Lado is noted for the point of departure for the former, Oller being critical of this, is noted for the latter (Baker, 1989, Morrow, 1981). What follows (below) takes us into the communicative legacy, a period or stage described as 'the Promised Land' by Morrow (1981).

3.2.1.5. Non-unitary competence hypotheses - (two)

Having arrived at the conclusion that the strong form of the UCH is untenable (Skehan, 1988: 213), that is, dissatisfied with the view that a single global language proficiency could account for communicative competence (Vollmer, 1983: 8), we observe a return to a multi-dimensional approach to the same issue. In this regard Vollmer (1983: 7-8) argues:

With the introduction of the 'communicative competence' notion and the beginning of functional testing we find that assumptions concerning the nature and structure of foreign language ability ... are now for the first time being stated more or less explicitly.

Skehan (1988: 213) and Palmer & Bachman (1981: 144) also hint at some consensus about the empirical support that a skill-based division of language proficiency has.

Writing about multi-dimensional models, Skehan (ibid.) identifies descriptive and working models. "Descriptive models outline a framework for, and the components of, communicative competence, while working models, in addition, indicate how the components inter-relate."

An overview of these models shows communicative competence with different dimensions or as having more than one component in the theoretical frameworks proposed by different authorities, for example, Canale and Swain (1980) Canale (1983) Bachman and Palmer (1984), Bachman (1986) (cited in Skehan, 1988: 213). While Canale proposes four components, i.e., linguistic, sociolinguistic, discourse and strategic, Bachman's model, according to Skehan, includes three trait factors, i.e. language and strategic competencies as well as skill and method factors (pp. 213-214). It can be seen that there is some rearrangement of the components in the latter. For instance, the fact that skill and method factors are added shows that not only competence, but competence and performance are dealt with (p. 214).

The model below is Bachman's (1990: 85) theoretical framework of 'communicative language ability' (CLA) which illustrates the interactions of the components with the language use context and language user's knowledge structures.

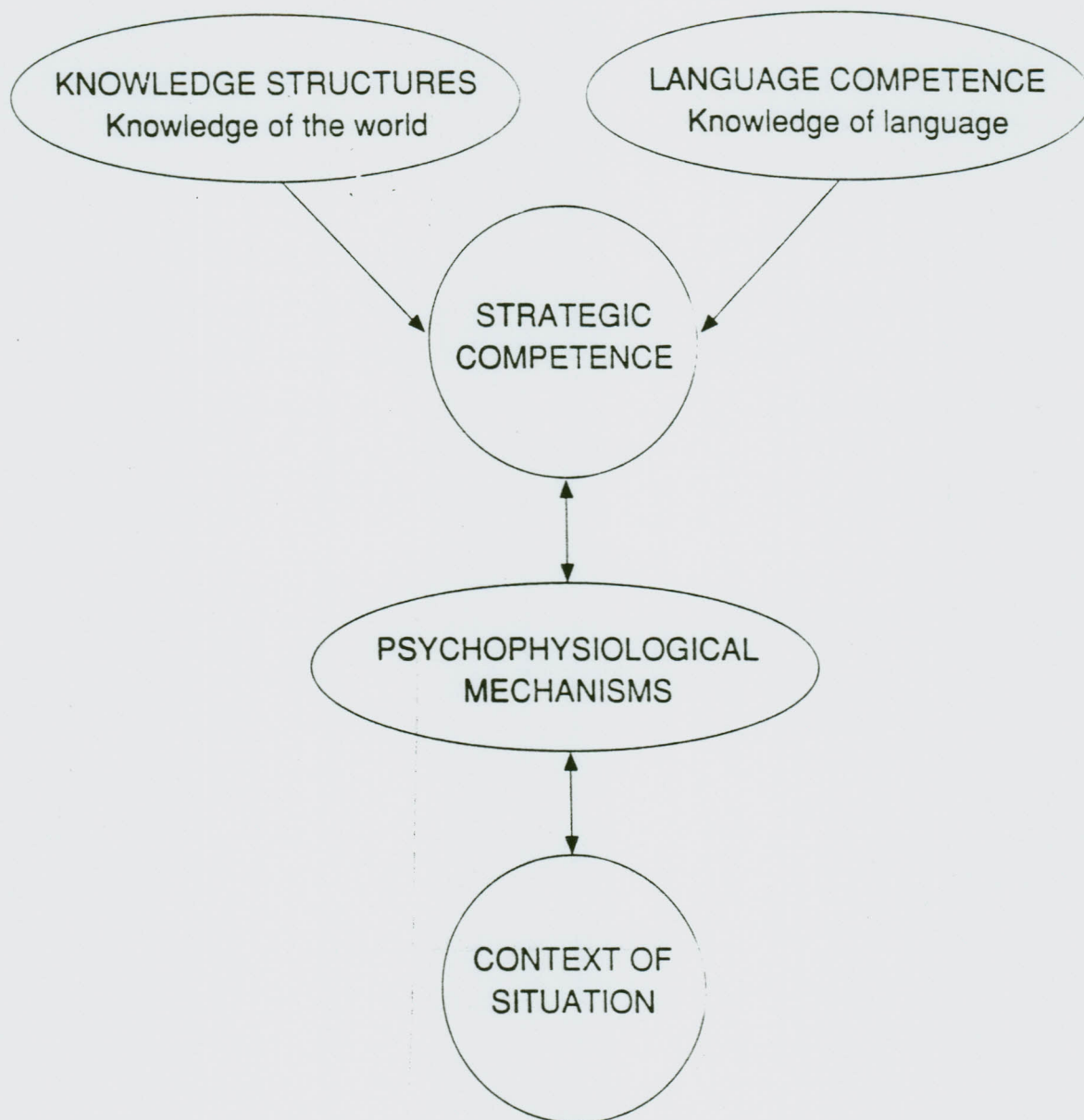


Figure 4.1 Components of communicative language ability in communicative language use

Strategic competence is given this central position because it is understood as the capacity that relates language competence to the language user's knowledge structures and the features of the context in which communication takes place. It is subdivided into three functions: assessment, planning, and execution. The psychophysiological mechanisms include the channel (auditory, visual) and the mode (receptive, productive) (pp. 107-108).

The components of language competence are illustrated separately in a 'tree' diagram (Bachman, 1990: 87):

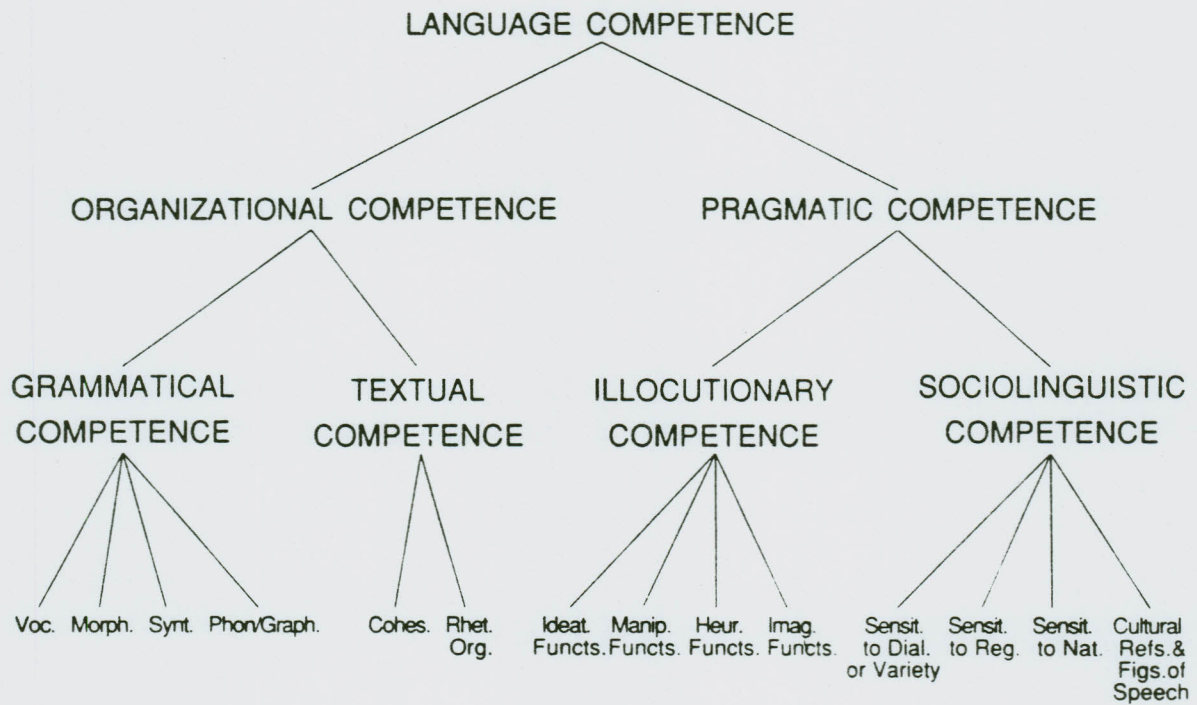


Figure 4.2 *Components of language competence*

Cummins (1984: 12) has also forwarded a two-dimension model, though his concern has been accounted for in the Bachman's model.

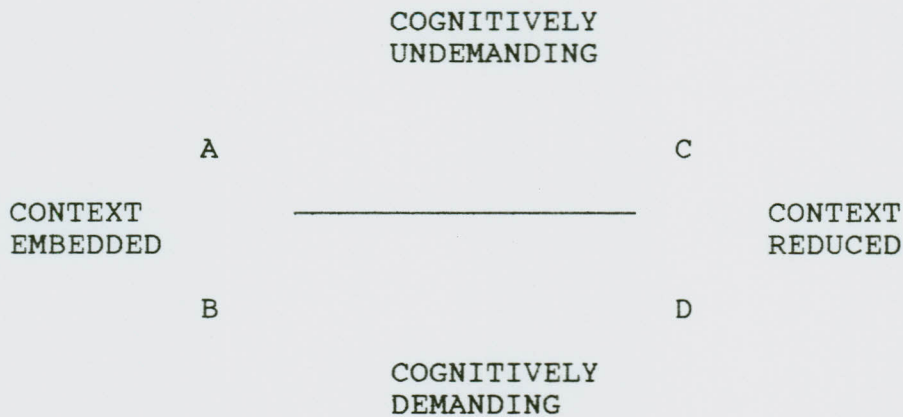


Figure 3: Range of contextual support and degree of cognitive involvement in communicative activities.

Cummins (1984: 13) explains:

It is important to emphasize that this is a continuum and not a dichotomy. Thus, examples of communicative behaviours going from left to right along the continuum might be: engaging in a discussion, writing a letter to a close friend, writing (or reading) an academic article. Clearly, context-embedded communication is more typical of the everyday world outside the classroom, whereas many of the linguistic demands of the classroom reflect communication which is closer to the context-reduced end of the continuum.

The linguistic tools for the tasks and activities lying in the upper end of the vertical continuum are so much mastered that little active cognitive involvement is needed for appropriate performance, whereas the contrary is true for those lying in the lower end (ibid.). He says, "The vertical continuum is

intended to address the developmental aspects of communicative proficiency in terms of the degree of active cognitive involvement in the task or activity (ibid.). Cummins (cited in Skehan, 1988: 214) argues that certain generalisations about language proficiency are invalid because:

... many generalisations about language proficiency are from context-embedded cognitively undemanding test situations, even though learners will have to deal with context-disembodied, cognitively demanding language use.

More recently, the claim that the structure of language proficiency consists of a general and some specific factors has been forwarded; e.g. Carroll (1983), Milanovic (1988) (quoted in Skehan, 1988: 212-213). Approaches to the nature of language proficiency are considered in the light of proficiency levels, too. For example, Sang et al. (ibid.: p. 214) propose elementary (e.g. basic elements of knowledge-pronunciation, spelling, lexis); complex (e.g. interaction of basic knowledge elements-grammar, reading comprehension); and communicative (e.g. interactive use of language). Still Canale (ibid., pp. 214-215) proposes a model of levels, not unlike Cummin's representation of competence, of a basic language proficiency (concerned with biological universals); communicative language proficiency (focusing on sociolinguistic and the management of conversations); and autonomous language proficiency (less-situation bound and use of language as a tool for thought).

There is evidence, from the reports of Sang et.al. (1986) and Milanovic (1988) (cited in Skehan, 1988), for instance, that lower levels of proficiency have a less integrated structure

whereas higher levels of proficiency have a more integrated structure, with the various skills converging together. This is so, Skehan (1988: 213) suggests:

... either because of the cumulative effects of instruction itself, or because of the inherent qualities of acquisitional processes, or perhaps because of the way different abilities can compensate for one another at higher levels.

Bachman's and Cummin's models are reproduced in this work because of their potential to explain the theoretical framework that underlies the tests developed in this study.

3.2.2. Purpose oriented language testing

3.2.2.1. Communicative language testing

Developments have also taken place in tests in the more specific areas. However, the achievement of performance conditions as proposed by Morrow and reformulated by Canale, when relating a model to actual performance, poses an important difficulty in communicative language testing. Skehan (1988: 215) enumerates them thus:

Genuine communication is interaction-based ...; unpredictable and creative ...; is situated in a context ...; has a purpose ...; is based on real psychological conditions ...; and is outcome evaluated, in that successful performance is judged in terms of whether communicative purposes have been achieved.

In order to achieve these conditions, the solution proposed is to use 'performance' tests that simulate some real-world target behaviour. Wesche (1987: 29) proposes three kinds of these tests, that is, direct assessment tests - where the situation is not manipulated, work sample tests - where tasks

are manipulated, and simulation techniques - where the entire testing situation is contrived. The point is that since these tests mimic the target behaviour, they are likely to achieve the conditions. She also holds that:

... such tests are more appropriate at higher proficiency levels and that they are better able to cope with situations when subject matter knowledge is important and when there is a possibility of compensatory performance disguising an underlying skill (ibid.).

However, doubts arise, for example, as to the extent of generalisability from these tests and whether direct tests necessarily replicate the performance conditions of a specific task. With respect to these performance tests Skehan (1988: 218) poses his doubt saying, "The main question here is really whether such techniques are likely to give measures for individuals which are accurate reflections of how they would perform when confronted with real-life communication." But, this doubt is even stronger when we consider structure-based, discrete-point tests which are removed from real-life use contexts.

3.2.2.2. Testing English for academic purposes

Early approaches to ESP focused on registers of specific areas in English based on the linguistic properties of the language. Applications of discourse analysis were aimed at describing varieties of English usage in terms of the syntactic and lexical properties of the language. Although this had very little impact on ESP tests, the influence of work in needs

analysis was great because it attempted to identify the special areas of language use for specific learners. Due to this, measures of performance could be designed and related to the performances that need to be predicted (ibid.). Moreover, Skehan (ibid.) observes:

Such tests seemed to by-pass the competence problem and the need to use theory since the relationship between test performance based on needs analysis and the criterion behaviour was transparent and direct.

Weir (1983: 105) too traces a change of view in recent approaches to ESP. As he puts it:

They were more concerned with predicting the communicative demands to be made on the learner in performing a role or series of related roles, rather than with the linguistic structures of the language per se. It was no longer a matter of simply identifying the appropriate register, but more a case of investigating the characteristic interactions in which language users were engaged, where they had actively to process discourse in the spoken or written medium and participate in real communicative behaviour.

Thus, ESP tests are geared towards establishing whether the individual has adequate communicative ability in a given context. In other words, ESP tests are concerned with proficiency in specific context, that is, not with general achievement but with specific skills required in this context.

Regarding proficiency in academic contexts, McEldowney (1976: 5) writes:

... neither the conversational nor idiomatic English that is required for

successful social intercourse, nor the type of English found in literature, is a central need for successful English medium study. What is considered to be central is a proficiency in the more expository, neutral, transactional type of English that is the medium of education.

Hence, there is need for needs analysis of the tasks and the skills (receptive or productive) required to perform in the appropriate situations. The application of needs analysis should, then, result in a specification of the general communicative tasks the candidate is to face in his academic context.

The results of needs analysis, however, tend to display numerous types of performances and situations with risks of going over-specific. In such circumstances, Weir (1983: 110) hints at the need to focus on representative abilities, and the more general and common tasks entailed by the various study modes across academic disciplines. This suggestion arises from the following finding:

In our investigations of the language events and activities ... students have to deal within ... academic environment and the difficulties they encounter therein, we discovered much that was common between students of different disciplines and at different levels.

Weir (1983: 59) points out in his study that:

... it is the problems caused by the 'higher order' language skills at the coherence and cohesion levels, together with study skills and study attitudes and habits, which receive most attention in the English for academic purposes (EAP) literature.

Concerning text content his report offers no conclusive evidence as to whether candidates were disadvantaged by writing tests with texts on topics outside their own subject areas. However, other sources suggest that test performance can be affected by background discipline. For example, Alderson and Urquhart (1984: 9-13), who conducted three studies to investigate the effect of students' background discipline, conclude that "Taken together, the studies do seem to confirm the importance of background discipline"; and based on one of the studies, they state, "The evidence is that background knowledge has significant effect."

3.3. Some basic considerations in testing

The design and construction of tests intended to assess language as communication, or more specifically, communicative type of ESP tests such as EAP are required to be needs related in terms of the tasks and skills required of the target candidate in view of the ultimate purpose. However, the question of generalisability and validity arise here.

A test may contain a set of performance tasks, which may in turn be slices of real-life communication, assuming that this is possible to achieve. But this is a very limited instance of the multitudes of activities and tasks in the multitudes of situations the candidate might find himself in within the specified context. Therefore, the test as a sampling instrument, is expected to furnish evidence to allow generalisation about the examinee's communicative ability in

the varieties of real-life situations. The data from the test should lend itself for the making of further inferences about the candidate's performances in the many possible situations he will have to face. Weir (1983: 92) expresses his experience thus:

The communicative type of EAP test that seemed most suited to our purposes implied the specification of performance tasks closely related to the learner's practical activities, that is, to the communicative contexts of situations he would find himself in; therefore we were faced with the problem of the generalisability of the tasks we selected.

Moreover, the desire to include a larger sample of tasks and more realistic test items results in a lengthy communicative test, which may mean that test efficiency or practicality is sacrificed to some extent (ibid.).

The selection of tasks is also related to test reliability. Kelly (1978: 226) elaborates:

That is, of all that a student is expected to know and/or do as a result of his course of study (in an achievement test) or that the position requires (in the case of a proficiency test), a test measures students only on a selected sample. The reliability of a test in this conception is the extent to which the score on the test is a stable indication of candidates' ability in relation to the wider universe of knowledge, performances, etc., that are of interest.

That is to say, test reliability is associated to test validity; and this brings us to what is called reliability-validity tension. Associated with sampling is test format or

type selected, with respect to the scoring entailed thereof, which also adds to the tension. In other words, communicative test types are best scored by criterion referencing, which involves an element of subjectivity as a factor that weakens reliability. Conversely, the choice of high reliability tends to undermine validity. But, unlike norm-referenced, criterion-referenced assessment has the advantage of revealing directly just what the candidate is able to do or not. Concerning the tension between these test qualities, Weir (1983: 77) has this to say:

This tension exists in the sense that it is sometimes essential to sacrifice a degree of reliability in order to enhance validity. If, however, validity is lost to increase reliability we finish up with a test which is a reliable measure of something other than what we wish to measure. The two concepts are, in certain circumstances mutually exclusive but if a choice has to be made, validity "after all, is the more important" (from Guilford, 1965: 481).

It is important, in its various aspects, because it indicates the suitability of the test as an instrument in measuring the domain we primarily wish to measure. Once this is confirmed, we turn to the consistency of the instrument.

Having considered issues of validity, reliability and efficiency (practicality) in terms of the 'tension' (Weir, 1988) that exists and the 'complementary' (Bachman, 1990) relationship they have between and among them in communicative type of EAP tests, we now look at each of the concepts in their own right.

3.3.1. Test validity

According to Henning (1987: 89), "validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is purported to measure. A test is said to be valid to the extent that it measures what it is supposed to measure."

However, a different view of the concept of validity has appeared during the past few decades. In this view, "validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores..." (Messick, 1988: 13). Furthermore, Bachman (1990) writes:

It has been traditional to classify validity into different types, such as content, criterion, and construct validity. However, measurement specialists have come to view these as aspects of a unitary concept of validity that subsumes all of them (P.236)

The different types of validity, according to the same writer, "...can be more appropriately viewed as complementary types of evidence that must be gathered in the process of validation" (p.243).

In this work, we proceed from the non-empirical to the empirical kinds of validity. First, the face and content validities, which require no use of formulae and have no coefficients or mathematical computations involved (Henning,

1987: 94) are presented. This is followed by construct validity, which is empirical in nature though it does not have any one particular validity coefficient (Henning, 1987: 98). It deserves this middle position to reflect the extent of empirical investigation it requires, and the comprehensive nature of the concept, i.e., its overlap with content validity, for instance.

Finally, criterion-related validities, i.e., concurrent and predictive validities are treated. These are empirical in that they involve the use of mathematical formulae for the computation of validity coefficients (Henning, 1987: 94).

3.3.1.1. Face and Content validity

Face validity is a non-empirical kind of validity concerned with the appearance of a test. Davies (1990: 23) writes, "Face validity concerns the appeal of the test to the lay judgement, typically that of the candidate, the candidate's family, members of the public and so on." Its role in public relations is considered important because "... testing is a public, political type of activity, an acting upon the world" (Davies 1990: 23). Still, there is a shadow of doubt regarding its importance in relation to the other validities; but given the public nature of testing, Bachman (1990: 288) states that:

... even those who have argued against 'test appeal' as an aspect of validity have at the same time recognized that test appearance has a considerable effect

on the acceptability of tests to both test takers and test users.

The other non-empirical kind of validity is also defined by Davies (1990: 23):

Content validity ... is a professional judgement, that of the teacher or tester. They rely on their knowledge of the language to judge to what extent the test provides a satisfactory sample of the syllabus, whether real (for achievement testing) imagined (for proficiency testing) or of the theory or model (for aptitude testing).

In this work, the views of candidates and language teachers was sought to see to what extent the tests satisfied their expectations of language tests.

3.3.1.2. Construct validity

Hughes (1989: 27) states how construct validity can be demonstrated by saying:

If the coefficients between scores on the same construct are consistently higher than those between scores on different constructs, then we have evidence that we are indeed measuring separate and identifiable constructs.

In this work, very little attempt is made to test separate language skills; rather, an attempt to integrate the different skills is sought.

For example, the written test (WT1) involves reading and writing; and the other two (LT1 & OT1) were designed to involve more than just listening and/or speaking. The three

new tests and the two external tests are compared with each other to provide some idea of the construct validity of each of the tests.

3.3.1.3. Concurrent validity

This is concerned with the validation of tests against some criterion measure of performance. "Another approach to test validity is to see how far results on the test agree with those provided by some independent and highly dependable assessment of the candidates ability, "writes Hughes (1989: 23). And for Henning (1987: 96), "It is criterion-related in the sense that the validity coefficient derived represents the strength of relationship with some external criterion measure." Thus, when a strong relationship or a high level of agreement between tests and criterion measures is observed, we can consider this as indicative of the validity of the new tests.

In this study, the new tests are compared with two external tests, Freshman English grades and ESLCE English grades to examine the extent of agreement in what they yield.

3.3.1.4. Predictive validity

This "___ is usually reported in the form of a correlation coefficient with some measure of success in the field or subject of interest," says Henning (1987: 97). In this procedure, test scores are correlated with some future

criterion of performance to find out to what extent the test(s) can predict candidates' future performance (Weir, 1988: 28; Hughes, 1989: 25).

In terms of the coefficients derived, Hughes (1989: 25) and Kline (1986: 5) point out that we can only expect a moderate one - something around 0.4 is generally considered satisfactory. Also, Downie and Heath (1974: 244) discussing criterion-related validity and the coefficients derived write that:

These validity coefficients tend to be much lower than reliability coefficients. An examination of the research over the years will show that they tend to fall within the band of 0.4-0.6, with a median value of about 0.5.

Although one can question the value of predictive validity because of the many variables involved, it is nevertheless, investigated herein as a matter of standard procedure.

In this study, ESLCE Grade Point Averages - which candidates already have, and University Semester Grade Point Averages - which would be obtained at the end of the semester, are used as criteria measures of performance.

The ESLCE Grade Point Average is used though it is not a future criterion of performance. This average grade is derived from scores on different subjects taught in high schools, and results of performance on this national examination, particularly English and Mathematics, are used for purposes of

employment and entrance to colleges and universities in Ethiopia. So, information on how the newly designed tests compare with this examination is believed to be interesting.

3.3.2. Test reliability

The other fundamental criterion for judging a test is its reliability. For Weir (1988: 34) "The concern here is with how far can we depend on the results that a test produces or ... could the results be produced consistently," and for Bachman (1990: 160) it "___ is concerned with answering the questions, 'How much of an individual's test performance is due to measurement error, or to factors other than the language ability we want to measure?'" . Thus, he offers a definition of reliability " as the proportion of the observed score variance that is true score variance" (1990:170).

Generally, "Reliability is thus a measure of accuracy, consistency, dependability, or fairness of scores resulting from administration of a particular examination." (Henning, 1987: 74).

While 0.9 or above is often mentioned as an appropriate coefficient of reliability for well made standardized tests, a coefficient of 0.7 (Baker 1989: 61; Kline, 1986: 3, for example) is hinted at as the minimum value. Since "Internal consistency coefficients are very suitable for use in computing the reliability of academic tests," (Downie and

Heath, 1974: 239), the Kuder-Richardson Formula 20 (KR20) was used for this purpose.

3.3.3. Test efficiency

This is concerned with the possible constraints involved in the implementation of the test, i.e., from its construction right up to the interpretation of the results and the making of decisions. Weir (1988: 37) reminds us that "A valid and reliable test is of little use if it does not prove to be a practical one. This involves questions of economy, ease of administration, scoring, and interpretation of results."

The human and material resources available need to be taken into account when designing the test. For example, the time available for construction, administration, scoring, and the quality and quantity of personnel and equipment needed are all crucial in terms of the overall cost. This is especially so because communicative tests require more resources, and demand the training of examiners and/or markers for the more complex marking involved.

Focusing on the length (duration) of the test which may cause fatigue effect on the testees, Weir (1988: 37) suggests that "It is thus highly desirable to make the test as short as possible, consistent with the need to meet the validity and reliability criteria referred to above."

Generally, efficiency, in the sense of financial viability, poses as problematic in the development of communicative tests for large-scale testing situations. But Weir (1988: 38) asserts that:

However problematic, 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 efficiency in the assessment of communicative skills.

In this work, the desire and attempt to achieve efficiency proved to be really problematic due to the extremely limited resources available.

3.4. General remarks

The development of a test is a process of making decisions, and of implementing these decisions at several stages. Roughly, a theoretical basis is necessary for what the test is measuring; a specification of the tasks and skills has to be drawn up to provide a good coverage of the theoretical basis; suitable items have to be written; it should be tried out and adjusted, where possible, before final administration; and finally, it should be validated against some criteria.

In terms of theoretical framework, the move is away from psychometric preoccupation to a more comprehensive, applied linguistic approach in testing (Vollmer, 1983: 6). Presently, the nature of proficiency is generally understood as partially divisible. Furthermore, the more purpose oriented communicative type of tests are favoured; and in addition, the

test format needs to achieve a certain balance between validity, reliability and efficiency as a measuring instrument (Weir, 1983: 79). The tester, too, has to consider the backwash effect and be alert against method effect.

The tester must strive towards the ideal, however distant it may appear to be. This is suggested by Alderson's comment regarding the problem of extrapolation (1981: 59).

... it may be that the issue of extrapolation is not (Yet) of crucial importance: even if we cannot generalise from performance in one situation to performance in a variety of situations, if we can say something about performance in one situation, then we have made progress, and if we can say something important about performance in the target situation so much the better. Ultimately the student will have to perform, despite the statistical evidence of the relationship between predictor and predicted, or the theoretised relationship between competence and performance [Author's emphasis].

Finally, Davies (1988) describes a communicative test thus: "It is likely to be more integrative and less discrete-point; more direct and less indirect; more criterion referenced and less norm referenced" (p.6). However, given the circumstances we are in in terms of the human and material resources at our disposal - if we come up with a test, for instance, that is "... largely integrative but indirect and norm referenced, or criterion referenced but discrete-point and indirect; and so on "(Davies, 1988: 6), then we are facing and progressing in the right direction.

Hence, the attempt, herein, is to design and validate a test of English proficiency for academic purposes, bearing some qualities of a communicative test.

The aim of this study is to establish the language demands of Addis Ababa University students and, based on this, to develop an appropriate test of English for first year students, including an investigation into the kind of information it would yield.

The methods employed in the design, development, and assessment of the claims of the new test of English for tertiary education in an Ethiopian context are described. In other words, information pertaining to the instruments used, and the procedures followed in their construction and in the collection of the data as well as the analysis of the data is presented in this chapter.

4.1. The subjects

The target population for this project are first year students at Addis Ababa University who take freshman English courses. The original intention was to include second year students taking Sophomore English. But because the course was offered during two different semesters for Social Science and non-Social Sciences, the idea was dropped because one would find either only Social Science or Science students at any one semester. Administering the tests twice was not considered wise.

The idea to pool students systematically from each section or class also posed difficulty. First either too many or too few volunteered to sign up for the tests from each class. Second, as witnessed during the pilot administration, those who signed up to take the tests failed to turn up for them. Moreover, the change of government and the kind of tension or insecurity that prevailed for a short while could have been the cause of the two obstacles cited above. This state of affairs resulted in the temporary closure of the university and irregularities like crash programmes when re-opened. This meant postponing schedules for the pilot and final administrations of the tests when re-opened, it meant administering the tests at the earliest time possible, and working under constraints like short semesters and heavy course work loads both for the students and the teachers.

The desire and attempt to find fund to pay the subjects or those students who are selected to take the tests proved futile too.

4.2. Instruments of data collection

The data analyzed in this study is collected using tests and questionnaires. There is a set of three newly designed tests and two external (international) tests as well as a set of questionnaires for the candidates and English language teachers.

4.2.1 The tests

4.2.1.1. The local tests

The new tests (Appendix 4-10) include the written test of reading and writing (WT1), the test of listening (LT1), and the oral test (OT1). These locally designed and constructed tests are a communicative type of English for academic purposes (See Chapter five for details). That is, they are skills-based, goal-oriented and based on needs analyses for their contents and objectives. In terms of format, some attempt is made to avoid complete unfamiliarity. For instance, some of the familiar formats like matching, fill in the blank, true or false, multiple-choice and its variants are retained though kept to the minimum. The 'Grammar' section, as a linguistic system, is also included for the same reason because our candidates are used to taking tests with a heavy dosage of structure. A further reason to retain this component is that students would not be reluctant to study grammar if they see that it is tested.

Some of the rather new formats in our context include cloze, editing, dialogue, partial dictation, labelling - based on a paragraph read, writing - based on a given diagram, filling in or completing blanks and supplying short answers based on various (authentic) sources. Since taking a listening test or an oral test in itself is unusual to our candidates, extra care was taken in the selection of test techniques. Partial dictation, for example, can be considered less unfamiliar

because our students are likely to have taken down dictations in other courses. Moreover, partial dictation and cloze are used as measures of general proficiency, and the scores on these would serve as core score when reporting overall performance on the test.

The candidates are required to record their answers on the question papers of the written test (WT1) and the test of listening (LT1). For the oral test (OT1), they are provided with source materials together with the instructions for the tasks they have to perform in pairs, i.e., with an examiner and with a fellow student.

4.2.1.2. The external tests

The external written test (WT2) is from a published TOEFL (1989), without the listening section, whereas the external listening test (LT2) is from Cambridge First Certificate in English (1986) (Appendix 12-17). This combination was decided on because the listening texts of the latter were accessible on recording. Fresh (local) recording of the printed listening texts of the TOEFL was not considered wise to do.

The TOEFL has two sections: structure and written expression, and vocabulary and reading comprehension. The candidates are required to read sentences and short passages and to answer the questions by choosing from the four options given. They blacken a circle of the letter of their choice on answer sheets they are provided with. While this test integrates

grammar, reading comprehension and vocabulary, the writing skill is extremely controlled, especially when recording answers.

The (Cambridge) First Certificate Listening Test has four parts. It is composed of a radio discussion, a radio news bulletin, a job interview, and a person talking about his job - all on recording. The candidates are required to listen to the recordings and answer the questions by putting a tick () in two or four box choices. Only four questions of part three require the candidates to write a one-or two-word answer. They are provided with a booklet containing the questions and spaces for recording answers. Generally, this test is the least integrated because the other skills are quite controlled in its design.

Both of these external tests (WT2 & LT2) are used to assess the English language mastery of students wishing to pursue their education in the U.S.A. and Britain from all over the world. One reason for their inclusion in the series is that they are approximately the same in purpose as the newly designed tests. That is, the purpose of the two external tests is to screen candidates wishing to enter educational institutions. Besides, some Ethiopian students who have completed high school and aspire to further education abroad are quite likely to take these tests. One can safely assume that English for academic purposes is assessed to some extent together with English for social survival, for example.

But, unfortunately, there is not statistical information available concerning these tests.

4.2.2. The questionnaires

There is also a set of questionnaires (Appendix 16-19) associated with the three new tests. These questionnaires, both for the students and language teachers, are designed in three parts to elicit information on the background of the respondents, and the face and content validities of the three newly designed tests. The candidates and the instructors were required to comment on about eleven statements for each of the three tests for the benefit of face validity. The questionnaire items for the students are about whether

- writing the test is enjoyable,
- the test measures language ability, and the grade will closely reflect this,
- it causes fear, anxiety or disappointment,
- the instructions are clear, and the layout is convenient,
- the test is different from English examinations they have taken,
- the time allotted is fair, and the length is reasonable,
- the passages and the sources of information are interesting, and simple enough for them.

The language teacher were also required to do the same on more or less the same statements. The slight difference is in

two items, of which one is a replacement and the other an addition. That is, teachers were asked whether the markability of the test is satisfactory, and whether the test (as a whole) is suitable for Freshman students.

In the third part of the questionnaire for instructors, the 19 items indicate the language skill(s) each section (with test item Nos specified) is meant to test. The language teachers were asked to show the extent to which they agreed or disagreed with these claims of the test. Again, in the third part of the questionnaire for students, 31 items of test sections (with test item Nos specified), passages and sources of information are listed. The candidates were asked to show their likes and dislikes by indicating whether each itemized part is good or bad. So, information gathered through these questionnaires makes up the subjects' biographical data which is used for sub-categorization of the sample. The reactions of the subjects as regards the quality of the tests is also used to assess the face and content validites of the said tests.

4.3. The construction of the instruments

The three newly constructed tests are based on two needs analyses conducted by B.L.Morris (1982) and Haile Michael Abera (1993). The content is drawn, and the aim and objectives of the tests are formulated from the needs established by these researchers.

Other sources like published textbooks and research undertakings of similar concerns were also consulted. The kind of language and/or study skills proposed and dealt with in published text-books for colleges or universities (at least beyond high school) were used as supporting expert opinions for the needs established at Addis Ababa University. Some examples of these include: James et al. (1979), Wallace (1980), Robinson (1981), Grellet et.al. (1982), Nuttal (1982), Swan and Walter (1985), and Soars (1987). Furthermore, the needs identified and the rank order of the macro-skills established at other places by other researchers were also consulted, and were found to more or less confirm the needs established here at Addis Ababa University. Typical examples include: Ostler (1980), Weir (1983 & 88), Hughes (1988) and James (1988).

During the construction of the instruments, both the tests and questionnaires were shown (given) to students and colleagues to react upon. Their opinions and reactions were used to moderate the tests and the questionnaires. This stage of trialling was found useful in tailoring the length of the test, detecting ambiguities in instructions, and in establishing answer keys and framing the marking guidelines, for example. The sometimes lengthy discussions, for instance, with colleagues over the tests clearly indicated the need for some kind of orientation or training for the listening and oral tests' examiners, and the marking of the paragraph writing in the written test.

4.4. Examining and marking

The written test (WT1) of reading and writing, and the test of listening (LT1) are constructed in such a way that they lend themselves to objective scoring with the exception of paragraph writing. Even here, an attempt is made to check the marker reliability of the markers on a sample of paragraph marking. Unlike these two, the oral test (OT1) is subjective or open-ended. However, three evaluators are used in order to achieve some degree of objective evaluation.

Marking guidelines are provided for the three tests including a five-point scale scoring sheet for the oral test (OT1) to be used by markers and examiners (Appendix 5,9,11). The examiners were also required to read aloud the listening text and the full text for the partial dictation during the listening test, and act as partners in two of the three tasks in the oral test.

While seven English language teachers took part at the pilot stage, 19 were involved at the final stage in the examining and marking of the tests. These language teachers were also required to comment on the tests by filling in questionnaires. In other words, they were asked to indicate the extent to which they agree or disagree with the claims of the tests regarding the tests' appearance and content.

These language teachers attended sessions of orientation, training and rehearsals of the roles they are to take in the examining and marking of the tests.

4.5. The data collection process

Initially it was proposed that the pilot administration would be conducted on 200 candidates and the final on 1200 candidates. The bargain over funds brought down these numbers to about 100 for the former and 350 for the latter.

So, language teachers were approached to convince their Freshman students to volunteer for the tests; and this was supported by notices put up at several places in each campus. As a result, 39 for the initial and 256 candidates for the final administration responded positively. These students were required to take the five tests, and immediately afterwards to fill in questionnaires about three of them.

Thus, the analysis is based on the reactions of 39 candidates for the initial administration, and 256 candidates for the final administration. But in the circumstances some second year students taking Sophomore English who turned up for the tests could not be refused.

Concerning the sampling of the subjects, a quick check of the sections or classes they have come from was rather reassuring for its spread, although a minority of the candidates did not bother to write their sections on the test papers or on the

attendances. For example, important subcategories like Science versus Social Science are represented. From such surveys the researcher is content that the sample reflects the target population.

Considering the candidates' background, we see that most of them attended government high schools outside Addis Ababa, and most of them are male students from the College of Social Science. Specifically, 211 went to government high schools, 33 went to private high schools, and six attended both at different times. An other six failed to give information. In terms of the location of the schools, 87 candidates come from schools in Addis Ababa, 158 come from schools outside Addis Ababa (in the provinces), and five had moved from one to an other. Six failed to provide information. Faculty wise, 167 are from the College of Social Science and 88 from the Faculty of Science. One student did not indicate his faculty. There are 236 male and 20 female students in the age range of 16 - 29 years.

As regards the sample size too, the researcher finds consolation in the minimum number suggested by Paul Kline (1986). Under reliability, Kline discusses the standard error of the correlation coefficient in relation to the size of the sample and writes:

Thus I recommend that reliability studies of tests should be carried out on samples of not less than 200, although larger samples are desirable. The k-R20 formula, which demands percentages of subjects putting the Keyed response, requires large sample for accuracy, and

200 here is certainly the minimum desirable (Kline, 1986: 123).

Sample size is also discussed under validity, especially criterion-related validity. For example, among the rules set out one reads: "Ensure that the samples are large enough to produce statistically reliable correlations which can bear factorial analysis. A minimum number for this is 200 "(Kline, 1986: 152).

So what is suggested is that generalisation about the tests is possible based on the data collected from the gross 256 candidates. Unfortunately, all the candidates did not sit for all the five or the three locally constructed tests, 207 candidates took the written test (WT1), 151 sat the listening test (LT1), and we were able to give the oral test (OT1) only to 62 candidates. For the international tests, we have 100 cases for external test A (WT2), and 151 for external test B (LT2). Thus, the sample size for most of the tests, except the written test (WT1) of course, falls short of the required minimum which calls for caution during analysis and interpretation of results.

4.6. Analysis of data

Appropriate statistics suited to small sample size (i.e., when there is the option) is used both in the initial and main analysis in order to gain information on the behaviour and quality of the tests.

4.6.1. Initial analysis

During the initial analysis (the pilot study), item analysis, descriptive statistics, and investigation into the reliability and validity of the tests are carried out (See section 5.4).

4.6.2. Main analysis

What is carried out during the initial analysis is repeated here with the aim of achieving a relatively greater degree of sophistication in the analysis. Some examples of such instances include comparison at finer levels of subgroups and subtests; and use of correlations to examine relationships as well as ANOVA to test the strength and interaction effects of the variables on test performance, etc.

4.6.2.1. Construct validity

The three locally constructed tests and the two external tests as well as their components are examined in relation to each other. This is done through correlations between the subtests of each of the five tests.

4.6.2.2. Criterion-related validity

While the criterion measures used for concurrent validity are scores on the two external tests, Freshman English and ESLCE English, those for predictive validity are ESLCE and University Grade Point Averages. This is carried out using correlations between the tests and the said criterion measures.

CHAPTER FIVE

THE DESIGN AND CONSTRUCTION OF THE TESTS

In this chapter, the aim is basically to draw up a specification of the English Language skills profile of Addis Ababa University students. But the procedures adopted to arrive at the specification and for its realisation are also described and discussed. The aim of the new test and the specific objectives as well as a detailed description of the subtests are offered.

5.1. Review of needs analysis

In this section, a review of needs analyses conducted to identify the communicative needs of Addis Ababa University students by Morris (1982) and Haile Michael (1993) are presented.

Morris's needs analysis involved instructors teaching courses other than language and students attending courses other than the language courses in eight faculties, excluding the Institute of language Studies and its departments. The faculty members were required to rank the four Macro-skills and four language activities within each macro-skill area in order of importance for their students in the courses they deliver. The students were also required to rate, in terms of the frequency with which they were engaged in, 25 language activities in the respective courses they were attending.

She then analyzed and established the most frequent and/or important language activities and proposed the language skills comprising them, using Munby's taxonomy of language skills.

Table 3: Most frequent language activities

(Morris, p.69)

Rank	ACTIVITY		MACRO-SKILL
1	PRACTICAL FACULTIES take notes from books/lectures	THEORETICAL FACULTIES	Writing, Reading, Listening
2	read/understand instructions	understand lectures	Listening, Reading
3	understand lectures	understand discussions	Listening
4	understand questions		Listening
5	understand spoken instructions		Listening
6		read instructions	Reading
7	read textbooks		Reading
8	read handouts		Reading
9	write short answers		Writing
10	use the dictionary		Reading
11	answer teachers' questions		Speaking
12	ask questions		Speaking
13	read other books (as source of information)		Reading
14	take essay examinations		Writing

For example, from the 25 language activities she started with she identified 13 (see Table 3) as the most important and/or frequent ones that Addis Ababa University students engage in during their academic years. That is, there are four for listening, four for reading, three for writing (two of which have the same language skills), and two language activities for speaking (see Table 4). Furthermore, she proposes about

104 language skills (see Appendix 2) that comprise these activities. That is, there are 24 for listening, 34 for reading, 25 for writing and 21 language skills for speaking. Of course, some recur for the different language activities and macro-skills. In this connection, Morris indicates that the receptive skills were ranked higher than the productive skills; and within each of these categories listening (slightly, in this case) and writing are in the lead.

Although Morris's needs analysis may be doubted for its freshness (relevance) for the 1990s target population, it has one obvious and attractive quality, ie, its cross-sectional sampling. In other words, the subjects (ie, the opinions of students and instructors) of the study are pooled from many (almost all) faculties and from first to final years. Thus, the needs identified are highly likely to be traced and plotted right through any undergraduate programme.

The needs analysis conducted by Haile Michael, available in the form of initial summary, is also considered to see if it would confirm or up-date the needs already established by Morris.

His study aims at establishing the language needs of Addis Ababa University students for the benefit of syllabus design by pooling information from the set of data gathered from different angles. His subjects involve 184 third year students and sixty subject teachers from Natural and Social

Sciences, namely, Natural Science, Technology, Social Science, Business and Economics, Education, and Law.

The students in Haile Michael's work, were required:

- to rate the four macro-skills according to their needs,
- to show their perception of abilities in the four macro-skills,
- to rate language subskills (these are language activities in Morris) according to importance,
- to show their perception of difficulties in the language subskills,
- to assess their abilities to follow courses taught in English,
- to assess the importance/relevance of the service courses,
- and to rate sources of information.

The subject teachers were also required to do the same except the last one.

The present study, which aims to develop a suitable test for Freshman English, focuses on the ratings of the four macro-skills and the language subskills from Haile Michael's data. This is so because information on these is more important and relevant for the drawing up of a specification for a Freshman language test. Furthermore, since both Morris and Haile Michael deal with these, it is direct and so convenient to compare the findings of the two needs analyses.

According to Haile Michael's data, the ratings of students and subject teachers in both streams rank the four macro-skills in the same way - that is, both agree that it is reading, listening (only slightly in the lead in Morris), writing and speaking, in descending order. Clearly, the receptive skills outrank the productive skills as in Morris's data. Considering the separate streams, Natural and Social Science Students, and Social Science subject teachers agree on the same order of importance except Natural Science subject teachers who rank listening first followed by reading. This rank order of the four macro-skills agrees perfectly with Morris's findings.

Regarding the language subskills (ie. language activities in Morris), there is complete agreement in the ranking of the language subskills associated with listening and speaking by students and subject teachers of both streams. These subskills are ranked the same by the separate streams for teachers and students, too. There is also a good deal of agreement in the ranking of the subskills associated with reading and writing.

Scrutinizing the slight disagreement, we notice that while "understanding exam questions" is ranked second by students and third by subject teachers, "reading reference books, journals" is ranked second by teachers and third by students, under reading. And under writing, concerning the subskills directly related to academics, "writing essay-type exam answers" is ranked second by students, but third by subject

teachers. The language subskills less directly related to academics, that is, writing "personal letters and business letters" are ranked differently in fourth and fifth place by these two categories of respondents. Thus, it can also be seen that the language subskills directly related to academics are ranked higher than the less academic ones.

Overall, most of the language subskills (language activities for Morris) dealt with by Haile Michael have been accommodated in Morris's study, or at least they have the opportunity to compete for inclusion in the specification. More specifically, in the receptive skills, what have been ranked one to three are clear candidates for inclusion in the specification. That is, "understanding lectures, following lectures to take notes, and following class discussion "under listening, and "reading textbooks, understanding exam questions, and reading reference books, journals "under reading are accounted for. Similarly, in the productive skills, "notes from lectures, books and essay-type exam answers "under writing, and "raising and answering questions "under speaking are also accounted for in her work. Therefore, we can safely say that this latter needs analysis confirms very closely the needs established by the former.

The language subskills of "presenting oral reports, and writing reports and research papers" in the productive skills of speaking and writing are left out because they appear to be advanced skills to be included in a Freshman test. Besides, they are skills intended to be developed in the Sophomore

English. Moreover, Haile Michael includes language subskills not directly related to academic but which may enhance study skills or promote language ability. These, however, are not considered for inclusion in the specification because of their indirectness. These include "understanding radio, TV, films, fluent speech with foreigners; reading fiction, magazines and newspapers, writing personal and business letters".

Finally, given that Haile Michael's sample population is third year students and their subject teachers in six faculties, and that the range of language subskills (four or five for each) for the four macro-skills is rather limited, language subskills like "presenting oral reports, speaking intelligibly, and writing reports and research papers" appear to have emerged at relatively high ranks. But these same language subskills have ranked low in Morris's work where the sample population ranged from second to final years in eight faculties, and the range of language activities (ie, 25) was much wider. The doubt is that these language subskills, in Haile Michael's work, may be true of just that sample, and not true of the general student body at large. Thus, they tend to be less attractive for inclusion in a Freshman test.

5.2. The specification

Given that the former investigator (Morris, 1982) has identified quite a number of language activities and proposed an even greater number of language skills, it is inevitable to find a way of rounding off the suggested size to a reasonable

minimum or a manageable size of language skills to be tested. That is, the idea is to establish a manageable set of skills, definitely less in size, but which can be satisfactorily argued to be representative and typical for the target students. The proposed skills that comprise the language activities are based on Munby's taxonomy of language skills, wherein these are treated discretely and in isolation. For this reason, the method sought to draw up the specification is one which is hoped to overcome this criticism.

The first approach, therefore, set out to consider for inclusion language skills proposed to be useful for several language activities in more than one macro-skill. In this procedure, first the language skills were plotted for each language activity and macro-skill. Then a summary of the frequency of language skills in terms of their occurrence in language activities and macro-skills was established. Once this was done, a series of 5 steps was followed in the reduction and/or inclusion process; and the process involves both mechanical and rational means either to include or exclude a language skill. The latter means took into account opinions of colleagues and information in published English textbooks.

Some of the proposed language skills (see Appendix 2) are dropped (crossed out) step by step as follows.

Step 1: Those language skills restricted to only one macro-skill are rooted out. But skill no 11 -

'Understanding intonation patterns..., skill no 12 -
'Producing intonation patterns..., skill no 19 -
'Deducing the meaning and use of unfamiliar lexical
items..., skill no 45 - 'Skimming to obtain the gist
and a general impression of the text', and skill no
46 - ' Scanning to locate specifically required
information...' are some of the exceptions. While
the first two represent the lower limit (ability)
for the receptive and productive skills
respectively, skill no 19 represents the vocabulary
component of the test. 45 and 46 are reading
specific (typical of it), and the latter poses as
essential for all the four language activities in
reading, ie., reading instructions, textbooks,
handouts, and using the dictionary.

Step 2: Those language skills restricted to only one
language activity in each of two or three macro-
skills are rotted out. Exceptions to this are skill
no 24 - 'Understanding conceptual meaning...' and
skill no 36 - 'Using indicators in discourse...'.
The former is believed to have a very high
likelihood of occurrence in the receptive skill, and
the latter pairs up with 35 (ie., 'Recognising
indicators in discourse...') in terms of receptive-
productive category of skills which offers the
advantage of representing both aspects of the skill.

Step 3: Those language skills that are unevenly shared (eg. in 2 vs 1 language activity) by two macro-skills are rooted out. The exception is skill no 44, ie, 'basic reference skills: understanding and use of...'. The attraction is that it could be assessed in both categories, giving the test an academic tint.

Step 4: Those language skills believed to be inappropriate to assess at the Freshman level are left aside. For example, skill no 32, 'understanding relations between parts of a text through grammatical cohesion devices ...' is left out. This is done because from experience former students were evaluated, and most of them were considered successful, based on predominantly structure-based examinations. Furthermore, the target population has gone and is going through syllabuses with a heavy overdose of grammar. Skill no 37, too, is dropped because it is developed and tested at the Sophomore level. This is so because it involves, for instance, the identifying of the main point through 'topic sentences in paragraphs'.

step 5: Adjustment is made in the distribution of language skills for the language activities and macro-skills in the light of proportion, exploitability of text, etc. Clear examples of this are skills 39 - 'Distinguishing the main idea from supporting

details...' and 52 - 'Transcoding information in speech/writing to diagrammatic display...', ie, the former is thought best to be assessed in the receptive where as the latter in the productive skills.

Table 4: Representation of skills in terms of language activities and macro-skills (after Morris, 1982)

LANGUAGE ACTIVITIES	LISTENING	READING	WRITING	SPEAKING
1. Listening to lectures	11 35 39			
2. Listening to instructions/explanations	11 24			
3. Listening to teachers' questions	11 26			
4. Listening to discussions	11 26 35 39			
1. Reading instructions		19 46		
2. Reading textbooks		35 39 44 45 46		
3. Reading handouts		51		
4. Using the dictionary		35 39 45 46 51 19 44 46 51		
1. Taking notes from lectures or books			44 51 52	
2. Short/essay examination answers			36 51 52	
1. Answering questions				12 36 51
2. Asking questions				12 36 51

This procedure reduces the proposed language skills to 19; and disregarding overlaps, to 13 skills in the receptive-productive areas. That is, in the receptive category, there are three language skills for listening and five for reading as well as two common ones for both; and in the productive category, there are two language skills for writing and one for speaking with two common to both. Generally, there are eight language skills for the receptive and three for the productive skills as well as two in common. We can view what has emerged differently; that is, there are two paired up skills (11 & 12, 35 & 36), some comprise both the receptive and productive categories (44, 51), and others are uniquely receptive-specific (11, 19, 24, 45, 46) or productive-specific (12, 36). So we see the likelihood of a language skill being assessed for both aspects, and both aspects being assessed in a language skill as well as each aspect (receptive/productive) being assessed in its own right. This is a desirable picture in that it appears to take into account the integrated nature of language skills.

Table 5: Final distribution of language skills for the macro-skills

Receptive			Productive			Total	
Macro-skills	No of skills	%	Macro-skills	No of skills	%	No of skills	%
Reading	7	37	Writing	(2) 4	21	(9) 11	58
Listening	*(3) 5	26	Speaking	(1) 3	16	(4) 8	42
Total	(10)12	63	Total	(3) 7	3	(13) 19	100

* Number of skills in brackets are excluding overlaps, ie., when the shared skills are included only in reading.

The distribution of language skills in Table 5 closely confirms Morris's findings in general. Both Morris's and Haile Michael's needs analyses are based on the opinions of students and content subject teachers. Accordingly, the table suggests that the written paper (ie, reading and writing) should receive greater emphasis than the aural/oral component of the test. Again the table reveals that the receptive skills should be emphasized more than the productive skills, ie, in the order of reading, listening, writing and speaking which is quite consistent with the needs analysis. This, therefore, is evidence of the fact that the procedure adopted herein to draw up the specification is a reasonably close reflection of the findings in the needs analysis.

This specification is also supported, at least to some extent, by a survey of skills treated in various published textbooks for learners of English.

The advantage of this approach is that it appears to permit a general (global) and a specific description of the candidates performance. Differently put, the strength of the individual can be described from his/her performance on the specific skill in a given macro-skill; and at least by implication, his/her overall potential in the other language activities or even in the other macro-skills could be detected. Considering skill No. 39, for example, the implications could be stretched across the language activities and the macro-skills. That is, given that the skill is tested in one mode of language activity and macro-skill, the candidate's ability can be

assessed for that particular activity and macro-skill; and his performance on this may indicate his potential ability in the other activities and skills. More specifically, if a candidate does well in "distinguishing the main idea from supporting details..." in reading textbooks, he/she might do the same in reading handouts; and even across macro-skills, he/she might do well in the same in 'listening to lectures', and/or show the distinction in 'taking notes and writing examinations' in the productive skill of writing. This assumption is, of course, based on the transferability of skills across modes of discourse. This assumption, of course, would be strengthened especially if the language skill is tested in at least two macro-skills. Moreover, it is also believed that skills may be integrated and that the distinction among the macro-skills is superficial, and this lends support to the approach.

Admittedly, this specification is drawn up using Munby's taxonomy of language skills. But, given the representation and distribution of the language skills for the language activities and the macro-skills in the specification, it can be seen that the assessment of them would be less discrete.

Finally, the following aim and objectives of the test can be offered from the specification.

The aim of the test is to assess a candidate's ability to function effectively in English for academic purposes, both in the receptive and productive skills.

1. Regarding the listening and reading skills, the objective is to assess the candidate's ability to understand spoken and written academic English, including reference sources.
2. Regarding the writing and speaking skills, the objective is to assess the candidate's ability to express inquiries and knowledge in speech and writing, including efficient note-taking.

5.3. Realisation of the specification

Below is a list of language skills for the four macro-skills (Munby, 1978).

Listening

Skill No.

- | | |
|----|--|
| 11 | Understanding intonation patterns ... |
| 24 | Understanding conceptual meaning ... |
| 26 | Understanding the communicative value (function) of sentences and utterances ... |
| 35 | Recognising indicators in discourse ... |
| 39 | Distinguishing the main idea from supporting details ... |

ReadingSkill No.

- 19 Deducing the meaning and use of unfamiliar lexical items
...
- 35 (see under Listening)
- 39 (see under Listening)
- 44 Basic reference skills: understanding and use of ...
- 45 Skimming to obtain the gist and a general impression of
the text
- 46 Scanning to locate specifically required information ...
- 51 Transcoding information presented in diagrammatic display
...

WritingSkill No.

- 36 Using indicators in discourse ...
- 44 (see under Reading)
- 51 (see under Reading)
- 52 Transcoding information in speech/writing to diagrammatic
display ...

SpeakingSkill No.

- 12 Producing intonation patterns ...
- 36 (see under Writing)
- 51 (see under Reading)

5.3.1. Description of the tests: content specification

The following specific objectives of the various subtests can be offered.

1. Reading Comprehension

Candidates are expected to demonstrate their ability in the following skills necessary for reading instructions, textbooks, handouts, and using the dictionary.

- 1.1 Skim a given passage for the gist or the main idea(s),
- 1.2 Scan a given passage for details or required specific information,
- 1.3 recognize the use and function of indicators in a written discourse,
- 1.4 deduce the meaning of unfamiliar words in a given context,
- 1.5 understand reference or source materials given,
- 1.6 understand tabular information instructions, and a form,
- 1.7 understand information presented in a diagram and in writing.

2. Writing

Candidates are expected to demonstrate their ability in the following skills necessary for taking notes from lectures or books, answering short and long (essay) examinations.

- 2.1 write notes of details or required specific information,
- 2.2 use reference or source materials,
- 2.3 use tabular information, instructions, and a form,
- 2.4 use discourse indicators,
- 2.5 transcode information presented in diagrammatic display into writing and vice versa.

3. Grammar

Candidates are required to show their awareness of grammatical accuracy by filling in gaps in a passage with selected deletion, and editing a paragraph with errors in grammar, spelling and punctuation.

4. Cloze

Candidates are required to draw on their entire knowledge of grammar, discourse structure, interaction between the text , knowledge of the world, etc. in order to fill in blanks in a passage with systematic deletion.

Cloze tests have the advantage of objective scoring, particularly when exact-word is used, and predictive validity. With regard to the latter, Brown (in Oller and Jonz, 1994:189) suggests that:

Perhaps of more interest to language teachers is the work on cloze as a measure of English as a second language...proficiency, which demonstrated substantial validity coefficients, .71 to .89, between cloze and standardized ESL proficiency tests.

Bachman (ibid., pp. 177-178) also confirms this when he writes:

There is now a considerable body of research providing sound evidence for the predictive validity of cloze test scores. Cloze tests have been found to be highly correlated with virtually every other type of language test, and with tests of nearly every language skill and component.

Thus, these qualities have been the attraction for the inclusion of the cloze component.

5. Listening Comprehension

Candidates are required to demonstrate their ability in the following skills necessary for listening to lectures, instructions or explanations, teachers' questions, and discussions.

- 5.1 understand spoken language,
- 5.2 distinguish the main idea(s) from supporting details,
- 5.3 understand conceptual meanings,
- 5.4 understand the communicative value of sentences,

5.5 recognise the use and function of indicators.

6. Partial Dictation

Candidates listen to a passage read out to them to show their understanding of it by writing in the missing words or phrase in the passage provided.

7. Speaking

Candidates are required to show their ability to respond appropriately with clarity and fluency of speech, and to ask and answer questions from sources of information.

Following is the table of specification which shows the method (format) employed in the construction of the tests. It also shows the various components (subtests) of the tests as well as the distribution of items (tasks) with the respective value (marks) for each.

Table 6: Table of specification

Written Test

Subtest	Section	Items	Value	Format
Reading Comprehension	A. Answering a question based on given passage	1	2	M/C
	B. Answering questions based on given passage	4	4	T/F
	C. Answering questions based on given passage	24	24	Short Answer & T/F
	K. Answering a question based on given paragraph	1	1	M/C
	L. Answering a question based on given diagram	1	1	M/C
	Total		31	32
Discourse Indicators	D. Answering questions based on given passage and statements	15	15	T/F

Subtest	Section	Items	Value	Format
	J. Using transitionals in a passage	13	13	Choose & fill in
	Total	28	28	
Vocabulary	E. Answering questions based on given passage	16	16	Matching
	Total	16	16	
Reference Skills	F. Answering questions based on given library catalogue cards	13	13	Short Answer
	G. Answering questions based on given dictionary page	10	10	Short Answer
	Total	23	23	
Writing	H. Filling in registration form based on given source and defined role	24	6	Fill-in
	K. Answering questions based on given paragraph	7	7	Labelling
	L. Writing a paragraph based on given diagram	1	5	Write
	Total	32	18	
Grammar	I. Filling in words in given passage (Rational cloze)	20	20	Choose & Fill-in
	M. Editing a given paragraph	10	10	Underline & Correct
	Total	30	30	
Cloze	N. Filling in words in a passage randomly deleted	40	40	Fill-in
	Total	40	40	
	Grand Total	200	187	
Listening Test				
Listening Comprehension	A. Answering a comprehension question on a passage read	1	1	M/C
	B. Answering comprehension questions on a passage read	8	8	Labelling
	C. Answering comprehension questions on a passage read	10	10	Choose & Write (short answer)
	D. Answering comprehension questions on a passage read and given statements	4	4	M/C or 2 options
	Total	23	23	
Partial Dictation	E. Filling in words or phrases from dictation on given passage of 279 words	20	20	Fill-in
	Total	20	20	
	Grand Total	43	43	
Oral test				
Speaking	A. Reading spoken language of 41 lines (1 page) long	1	5	Speak (dialogue) in pair
	B. Responding appropriately in a double dialogue of 11 lines (1/2 page) long	1	5	Choose & respond to partner
	C. Filling in missing information based on given source	2	5	Ask and answer Questions
	Total	3	15	
	Average		5	

5.3.2. Description of the subtests

What follows is further information on the various subtests or components of the test.

Written Test

READING COMPREHENSION

The candidates are required to read a passage (54 lines long) dealing with the progress and problems in the supply of power in Africa. This is followed by:

- 1 comprehension question of multiple-choice,
- 4 comprehension statements to be marked True or False.
- 16 items of comprehension requiring candidates to write short answers(words or phrases)in note forms,
- 8 comprehension statements (in a table) to be marked True or False,
- 15 statements made about transitional words to be marked True or False,
- 16 lexical items requiring candidates to match them with dictionary meanings.

REFERENCE SKILLS

A set of six library catalogue cards are provided for candidates to use as a source of information to carry out the following tasks:

- 1 item of alphabetizing the names of authors,

-7 items dealing with information on books.

Also candidates are provided with a dictionary page to use as a source or information to answer

-10 items of writing short answers.

WRITING

Candidates are presented with a page of course offering in undergraduate programme with a definition of their role (status) to fill in (complete) a registration form:

-24 blank spaces for the value of 6 marks.

They are also given a short (9 lines) paragraph on a description of aspects of agriculture, followed by:

-1 comprehension question of multiple-choice,

-7 items of labelling a diagram.

In addition, candidates are provided with a diagram on the classification of formal education in Ethiopia, followed by:

-1 comprehension question of multiple-choice,

-1 task of writing a complete paragraph.

GRAMMAR

In this section, the items involve verb forms (for number, the tenses and modals), prepositions, articles, comparatives, and relative pronouns.

Candidates are provided with a passage (30 lines) on Gross National Product (GNP)with some selected deletion of words, and a list of choices (words) at the top of the passage (page). They then choose and fill in

- 20 gaps with one word.

They are also provided with a paragraph on agricultural education, containing errors in grammar, spelling and punctuation. Here, they are required to underline and correct

- 10 errors.

CLOZE

Candidates are given a passage (44 lines) on population with words systematically deleted. About 5 lines of the first and final paragraphs are left intact. They are then required to read and fill in

- 40 gaps with exact words.

Listening Test

It was decided to use real, human voice for this test. This was done to avoid the lack of tape recorders and to approximate the reality as closely as possible. The reality, in the Addis Ababa University context, is that students would very often attend lectures delivered by Ethiopians predominantly, and by other nationals sometimes. Besides, the testing of the listening skill in itself and the use of tape recorders could, it was thought, prove unfamiliar to the candidates.

Thus English language teachers available at Addis Ababa university who were also willing to be involved were given orientation and training on the reading of the listening text and the partial dictation. In order to achieve a reasonable degree of uniformity, these teachers had to meet for sessions of discussion and rehearsal. The researcher had to explain and define the nature and the objective of the test and lead the sessions on issues like: who should read what, how many times should the texts be read, and at what speed or pace should they be read? etc.

The passage read out (3 1/2 pages long) is on how to use a library, and it was read twice by two instructors (examiners) in turns. This is followed by:

- 1 listening comprehension question of multiple-choice,

- 8 listening comprehension questions by labelling a diagram (library card),
- 10 listening comprehension questions where candidates are required to write short answers by selecting (choosing) from 12 choices given at the top of the page.

Also, statements (sentences) taken from the passage, containing transitionals are given with paraphrases for each. The candidates are then required to choose the paraphrase with the nearest meaning

- 4 items with 2 options each.

PARTIAL DICTATION

A full (complete) passage (279 words long) on taking notes from a lecture was read out twice by two instructors. The students, on the other hand, were provided with some passage with some selected words or phrases deleted. They were required to listen and fill in

- 20 gaps.

Oral Test

Given the amount of time and space required, the examiner judgement involved, and the cost involved in the special arrangements required for these, the oral skill is often

considered difficult to test, particularly in a large scale testing situation. But a modest attempt, comprising half indirect material, is made to assess this skill. Clear instances of this are the dialogues, particularly reading aloud the single dialogue.

Some writers like Rees (1980), Madsen (1983), Underhill (1987) and Shaw (1988) acknowledge that reading aloud appears in realistic situations in the real world. "Speeches, reports, minutes..." (Rees, 1980: 115) are read aloud in everyday life; reading aloud "...can produce connected speech...(Madsen, 1983: 149), which appears in everyday communication; and "...in real life, or at least in the education system, one does have to read things aloud. Many people need the support of the written word to be able to give a clear presentation to a group" (Shaw, 1988: 42). Moreover, Rees points out what is involved in reading aloud when he says:

What we are looking for is the expression of meaning by reading aloud intelligently, that is by paying attention to correct pausing, whether indicated by punctuation or not; by reading complete phrases, and not word for word; by adopting an appropriate speed and rhythm, and not distorting pitch, stress, and intonation; and reading in a relaxed manner...(1980: 119).

Madsen (1983) and Underhill (1987) say that reading aloud is suitable and useful to assess the skills of oral production. According to Underhill, reading aloud and its variants are suitable to assess the mechanical skills of language production; and he goes on to add that:

...as well as testing mechanical speaking skills, these techniques also call on the ability to add meaning at the sentence and discourse levels. This is a kind of fluency, although it is harder to mark; but it is usually clear to the assessor if a learner is able to pronounce each word correctly in its isolated form but is unable to put the words together to form a stream of meaningful speech (1987:77).

Concerning reading aloud dialogues in particular, he says:

Reading one part in a dialogue adds an element of interaction, of a limited kind; the student is speaking to, and with, another person in the foreign language, although the conversation is wholly predictable. (1987: 77)

When reading aloud is used to assess the speaking skill, "We can see how well our students have mastered the sound-symbol relationships in English. In some programs oral reading is important, and this test technique is obviously a good, direct way to measure oral reading. It can even provide a little help in checking fluency (Madsen, 1983: 154)

From experience, some students at Addis Ababa University read aloud when studying in groups in the corridors of classroom buildings, and a few others are sometimes employed to read aloud to the non-sighted student population. Moreover, students know that they are very likely to use this skill in the near future when they will be required to deliver oral presentations of essays, reports, projects, and etc. Reading aloud a dialogue, as an integrated test of the skills of reading, speaking, and listening was considered appropriate to include as a component in the oral test. So, it is believed

that candidates begin the oral test with a relatively easy and meaningful task, which is not too unfamiliar.

The evaluation, on a 5-point scale, was done by three instructors for every pair of candidates. They were given orientation and training, and marking guidelines with scoring sheets. Again teachers had to come to several sessions of discussion and rehearsal. The researcher, too, had to give a lot of explanation and definition on the nature and objective of the test. The issues discussed and rehearsed include definition of what is assessed, seating arrangements, sharing out roles, use of the scoring sheet, ways of leading in and putting candidates at ease, etc. All this was done to achieve objectivity in their evaluations of the candidates.

In the discussion of the assessment of oral production, Madsen (1983) cites scales with five and six levels as examples for a holistic scoring method and Underhill (1987) uses a five levels scale to demonstrate his definition of rating scales. In Underhill, simplicity seems to be the main point in his practical advice on the construction of a scale. In his own words:

The fewer levels you have, the easier it is to assess, and the higher the reliability will be. Keep it as simple as possible. A rating scale will only work well if the assessor can hold it in her mind while listening or talking to the learner, and does not have to keep referring to a large manual to tell her what to look for (p.100).

So a 5-point rating scale (below) is used to assess the oral production of the candidates. The five point is preferred to match the existing letter-grade system in the university, ie., A, B, C, D, and F. It was also attempted to keep the descriptors as short and clear as possible for the benefit of the examiners.

5 = Very comprehensible and appropriate, and is able to ask and answer questions very well.

4 = Reasonably comprehensible and appropriate, and is able to ask and answer questions reasonably well.

3 = Comprehensible and appropriate, and is able to ask and answer questions.

2 = Partially comprehensible and appropriate, and is rather poor in asking and answering questions.

1/0 = Incomprehensible and inappropriate, and is very poor in asking and answering questions.

(See Appendix 13 for the scale used for each of the three sections)

The oral test has three sections: a single dialogue between two Addis Ababa students when they first meet in a dormitory, a double dialogue - extending invitation and discussing a news item in the paper, and a class schedule with information gap.

In the latter, each partner has different bits of information missing which can be found with the other.

So, each candidate reads a dialogue (1 page long) with an examiner, responds appropriately to the examiner in the double dialogue, and asks and answers questions with the other partner based on the source (class schedule).

5.4. Initial administration

The design, construction and validation of a new test requires the trialling and piloting of the test on a sample before the final administration to the target population. At the trialling stage, opinions of colleagues and students have been taken into account to moderate the tests.

Hughes (1989: 52) states the need for piloting more explicitly.

Even after careful moderation, there are likely to be some problems with every test. It is obviously better if these problems can be identified before the test is administered to the group for which it is intended. The aim should be to administer it first to a group as similar as possible to the one for which it is really intended.

Broadly speaking, the results of the analysis of the initial administration on a sample (i.e., the piloting of the tests) yields information about the behaviour of the tests (Baker, 1989: 46). More specifically, the results gained from pre-testing could provide useful information regarding the

performance of the students (as individuals and as a group), and the performance of each of the items that make up the test (Heaton, 1975: 174; Madsen, 1983: 180).

Therefore, the aim here is to report the results of the analysis of the initial administration of the series of tests which were used for further moderation of them; i.e., the results of item analysis, and an investigation into the reliability and validity of the tests are reported.

The pilot study was conducted on 39 students taking the Freshman English course. These students were required to take the three newly designed tests and the two external tests, and immediately afterwards to fill in questionnaires about the first three. Seven English language teachers were also involved in the examining and marking of the tests. These language teachers were also required to comment on the tests by filling in questionnaires.

Considering the candidates' grades in ESLCE English and Freshman English 101A, it should be noted that they were high achievers in English examinations. We should note that 76, 21 and 3 per cent of them had grades of A's, B's and C's respectively in ESLCE English; and 28, 36 and another 36 per cent of them have A's, B's and C's respectively in Freshman English 101A. These letter grades are from two norm-referenced examinations.

5.4.1. Descriptive statistics and Test reliability

Table 7: Descriptive Statistics and Reliability coefficients (KR21)

Test	Max. Mark	No of items	Average	Standard deviation	Coefficient of Discrimination	Test Reliability
WT1	187	199	130.5	22.53	0.121	0.915
LT1	43	43	36.07	5.08	0.118	0.793
OT1	5	4	3.62	0.75	0.150	0.530
WT2	100	100	65.25	12.58	0.126	0.865
LT2	30	41	12.48	6.24	0.208	0.797

Considering the average score (in Table 7) of the tests, it can be seen that the candidates have found the three new tests and the first external test (WT2) rather easy, unlike the second external test of listening (LT2). Perhaps, this is not surprising given the fact that the sample population is rather homogeneous in language achievement levels as evidenced in their grades for the ESLCE English and Freshman English 101A.

It can also be seen, considering the standard deviations, that most of the tests spread out the candidates across the score range for each test rather closely. The coefficients of discrimination, which are expressions of the standard deviations as proportions of the total marks for the different tests, reveal that the second external test of listening (LT2), followed by the oral test (OT1), spread out the candidates more effectively than the others.

The particularly minimal difference between the coefficients of the new written test (WT1) and the first external test (WT2), and to some extent that between the new listening test (LT1) and the second external test (LT2) is clear evidence of the fact that the group is fairly homogeneous with a narrow range of proficiency levels.

Perhaps one could reasonably assume that, at least, the two external tests would have spread the candidates rather more widely if it had not been for the homogeneity of the group. Note the fact that the three new tests should behave more or less the same as the two external tests is quite encouraging.

The Kuder-Richardson Formula 21 (KR21) was used to compute the reliability of the tests. Thus, a quick glance down the reliability column in Table 7 shows that the tests had quite satisfactory coefficients, except the oral test (OT1) which was mainly subjective in nature.

5.4.2 Item analysis of the tests

Item analysis is a useful procedure for revealing information about the performance of the test items comprising a test. It allows us to examine all the items in terms of their level of difficulty, level of discrimination (Heaton, 1975: 173), and contribution to the total test (Hughes, 1989: 160).

The scripts of the candidates who took the series of the tests in the pilot study were rank ordered according to their total

score from the highest to the lowest. Applying suggestions for small samples (Harrison, Heaton, Madsen, Downie and Heath, for example) the top 1/3 and the bottom 1/3 of the group were used for the analysis.

Regarding facility values, any value falling between 0.4 and 0.6 is generally acceptable, 0.5 (50%) being the most desirable value (Harrison, 1983: 128 & 131; Heaton, 1975: 173). But other ranges are also suggested: for example, Kline (1986: 143) 0.2 to 0.8, Baker (1989: 54) 0.25 to 0.75, Heaton (1975: 173) 0.3 to 0.7, and Madsen (1983: 182) 0.3 to 0.9.

Discrimination indices, cited by Dejene (1990: 72), ranging from 0.2 to 0.8 are considered acceptable, with 0.67 as the most desirable value. But, while a value of 0.3 or higher is satisfactory for Baker (1989: 54) and Harrison (1983: 131), it is 0.15 or higher for Madsen (1983: 183).

For item-test /-total correlation, the satisfactory levels are set at 0.3 or above by Hughes (1989: 160) and beyond 0.2 by Kline (1986: 143).

Generally, given the relatively high achieving nature of the group, the more relaxed levels are kept at close range during selection, ie., between 0.2 and 0.9 for facility value, between 0.2 and 0.8 for discrimination index, and 0.2 or above for item-total correlation.

Each item of the 200-item written test (WT1), with the exception of one task of writing a paragraph, was scrutinized in the light of these three criteria. Especially in the cloze section and at other places in the written test, some attempt to change the items (questions) was made as well as rejecting those items that failed to satisfy the requirements of the criteria. First, items that met the three criteria were retained without any change, followed by the acceptance of those items that fulfilled any two of the three criteria with some changes made to most of them.

Overall, the items included in the revised written paper had a range of facility values of 0.18 to 0.91 (with no more than three items at the extremes), and discrimination values of 0.18 to 0.75. Thus, the moderated written test (WT1) has 124 items and/or tasks for further analysis at the final administration.

Similarly, the 43 items of the listening test (LT1) were examined. Accordingly, section D, which was on transitionals, was wholly rejected. Quite a bit of change was made particularly to the partial dictation, and slight changes elsewhere in the paper.

Generally, the items included in the revised listening test had a range of facility values of 0.39 to 0.89, and discrimination values of 0.22 to 0.78 (and one item of 1.00). Thus, the moderated version of the listening test has 30 items for further analysis at the final administration.

The oral test (OT1) is not found amenable to item analysis, and the whole set was retained for the final administration with only slight changes made to the instructions. This can also be justified given the positive views of the candidates and the examiners about this component of the test.

According to the item analysis, it can be said that the two newly designed tests (i.e., WT1 & LT1) are rather easy, and this may be due to the nature of the sample. The written test had an overall mean facility value of 0.71 and the listening test had 0.82. The overall mean discrimination for the former is 0.21 and for the latter 0.28. Notice, here, that the listening test is easier than the written test, but it also discriminates better.

Given the small size of the sample which, by coincidence, happens to be rather highly motivated with high language achievement levels, it was thought best to retain the low difficulty of the tests in many of the cases. This is done because the full target population may not be as highly motivated as the pilot group.

A quick analysis of the two external tests was also made to see how they have functioned, though not for revision. Accordingly, the first external test (WT2) had facility values ranging from 0.06 to 1.00, and discrimination levels ranging from -0.25 to 0.75. The second external test (LT2) had facility values ranging from 0.17 to 0.89, and discrimination levels ranging from -0.11 to 1.00. In terms of these two

criteria, 53 per cent of the items in the first external test and 78 per cent of the items in the second external test were found acceptable. The rest of the items in both tests were candidates either for rejection or for some revision. Comparing the two, it can be observed that the second external test of listening behaved much better than the other.

5.4.3. Test validity

5.4.3.1. Face and Content validity

Candidates who sat the tests and language teachers who were involved in the examining and marking of the tests were invited to give their views on them in questionnaires. Information obtained from these were used to judge the face and content validities of the tests. In other words, they were asked if the tests actually met their expectations of language tests, both in appearance and content. An attempt is also made to find out the likes and dislikes of candidates regarding the components of the tests in this study. A component, for example, judged 'very bad' by a majority of the candidates may need to be investigated further.

In reporting the responses to the questionnaires, the five-point scale is reduced to three categories: disagree, neutral and agree; or bad, neutral, good in the case of students. When reporting in percentages, any value greater than 33 per cent is considered substantially meaningful- both for the students and teachers.

Table 8: Students' response in frequencies and percentages

SQ1WP	SD	%	D	%	N	%	A	%	SA	%	NR	%	Tot	%
FVIS	5	1.4	20	5.5	40	11.0	168	46.3	124	34.2	6	1.6	363	100
CIS	12	1.7	26	3.8	108	15.6	327	47.2	185	26.7	35	5.0	693	100
Oval	17	1.6	46	4.4	148	14.0	495	46.9	309	29.3	41	3.8	1056	100
<u>SQ2L</u>														
FVIS	0	0	9	3.3	39	14.4	133	49.3	84	31.1	5	1.9	270	100
CIS	0	0	4	2.4	17	10.5	91	55.2	44	27.2	6	3.7	162	100
Oval	0	0	13	3.0	56	13.0	224	51.9	128	29.6	11	2.5	432	100
<u>SQ3S</u>														
FVIS	1	0.4	8	3.5	29	12.6	102	44.4	88	38.3	2	0.8	230	100
CIS	0	0	0	0	6	6.5	38	41.3	40	43.3	8	8.7	92	100
Oval	1	0.3	8	2.5	35	10.9	140	43.5	128	39.8	10	3.0	322	100

Abbreviations: SQ1WP, SQ2L, SQ3S = Student Questionnaire on the Written, Listening, Speaking tests respectively.
SD = Strongly Disagree; D = Disagree; N = Neutral;
A = Agree; SA = Strongly Agree; NR = No Response.
FVIS = Face Validity Items; CIS = Content Items;
Oval = Overall, ie, combination of the two.

Overall, about 76 per cent of the students have expressed their positive views regarding the appropriacy of the written test (WT1), and about 29 per cent showed their strong agreement with the claims of the test. More specifically, 81 per cent and 74 per cent thought that the test had good face and content appearance, respectively. 34 per cent expressed their strong agreement with the claims of face validity of the test, and 27 per cent thought the components of the test were very good.

About 82 per cent expressed a positive view about the quality of the listening test, and 30 per cent strongly agreed to this. More specifically, 80 per cent held positive views about the face validity, and 83 per cent liked the sections of the test. 31 per cent strongly agreed with the face validity, and 27 per cent thought the sections were very good.

About 83 per cent of the candidates felt the oral test was appropriate, and 40 per cent expressed strong positive views. 83 per cent believed the test had good face validity, and 85 per cent thought the tasks were good. 38 per cent strongly agreed to the former, and 44 per cent believed the latter were very good.

Table 9: Teachers' response in frequencies and percentages

LTQ1WP	S D	%	D	%	N	%	A	%	SA	%	NR	%	Tot	%
FVIS	1	1.3	6	7.8	10	12.9	33	42.9	25	32.5	2	2.6	77	100
CVIS	0	0	0	0	10	13.0	30	39.0	37	48.0	0	0	77	100
Oval	1	0.6	6	3.9	20	13.0	63	40.9	62	40.3	2	1.3	154	100
<u>LTQ2L</u>														
FVIS	0	0	0	0	8	11.43	36	51.43	26	37.14	0	0	70	100
CVIS	0	0	0	0	6	17.1	14	40.0	14	40.0	1	2.9	35	100
Oval	0	0	0	0	14	13.0	50	48.0	40	38.0	1	1.0	105	100
<u>LTQ3S</u>														
FVIS	0	0	2	3.33	16	26.67	33	55.0	9	15.0	0	0	60	100
CVIS	0	0	0	0	4	22.22	9	50.0	5	27.78	0	0	18	100
Oval	0	0	2	2.56	20	25.64	42	53.85	14	17.95	0	0	78	100

Abbreviations: LTQ1WP, LTQ2L, LTQ3S = Language teachers Questionnaire on the Written, Listening, Speaking tests respectively.
SD = Strongly Disagree; D = Disagree; N = Neutral;
A = Agree; SA = Strongly Agree; NR = No Response.
FVIS = Face Validity Items; CVIS = Content Validity Items;
Oval = Overall, ie, combination of the two.

Overall, 81 per cent of the language teachers agreed that the written test met their expectations, and 40 per cent expressed their positive views strongly. Specifically, 75 per cent agreed with the appearance, and 87 per cent agreed with the content of the test. And, 33 per cent held strong positive views about the face validity, and 48 per cent held the same strong views about the content of the test.

86 per cent of language teachers agreed that the listening test had face and content validities, and 38 per cent showed strong agreement regarding these qualities. Specifically, 89 per cent agreed with the claimed appearance, and 80 per cent agreed with the content of the test. 37 and 40 per cent respectively expressed strong positive opinions about the face and content of the test.

On the whole, 72 per cent of the respondents agreed with the claimed appropriateness of the oral test, and 18 per cent strongly agreed. 70 per cent had positive views about the face of the test, and 78 per cent agreed with the content of the test. 15 and 28 per cent of respondents expressed strong agreement regarding the face and the content of the test.

Therefore, given the information in Tables 8 and 9, one can say that the tests have quite acceptable face and content validities as measures of language ability.

5.4.3.2. Construct validity:

Table 10: Correlation coefficients between tests

WT1	LT1	LT2	WT2	
-	-	-	-	WT1
0.60+	-	-	-	LT1
0.29	0.50+	-	-	LT2
0.72+	0.75+	0.37	-	WT2
0.72+	0.55+	0.45+	0.64+	OT1

+ = Significant at the 5% level

Comparing the new written test (WT1) with the others, we observe that there is a meaningful overlap and a significant relation with one of the external tests (ie, WT2), and least overlap with the other external test of listening (LT2). This is evidence that the test as an integrated test of reading and writing is doing the job it is designed for.

There is also evidence that the new listening test is doing its job given the meaningful overlap and significant relation with the external test of listening. The relation becomes clear when we see that the correlation coefficient for the external test of listening is higher with the new listening test than it is with the other tests. That the new listening test should show higher correlation with the other tests is only a reflection of the fact that other skills are highly controlled in the external test of listening.

The new oral test, too, bears more meaningful overlaps and significant relationships with the other tests than with the external test of listening (LT2), which is the least integrated test. The meaningful overlaps and significant relationships between the oral test and the written tests could be due to the amount of reading involved in both, whereas, though to a lesser extent, that between the oral test and the listening tests may be due to the amount of listening involved in both-listening to partners in the oral test, for example. Again, that the oral test should correlate the least with the external test of listening is evidence that the test is doing its job as an integrated oral test.

5.4.3.3. Concurrent validity

Table 11: Correlation coefficients between the new tests and the criterion tests and grades

WT1	LT1	OT1	WLT1	WOT1	LOT1	WLOT1	
0.72+	0.75+	0.64+	0.78+	0.73+	0.78+	0.79+	WT2
0.29	0.50+	0.45+	0.36	0.22	0.53+	0.30	LT2
0.53+	0.71+	0.53+	0.64+	0.59+	0.69+	0.65+	FLEG
0.11	0.28	0.14	0.22	0.21	0.42	0.25	ESLEG
0.46+	0.68+	0.45+	0.60+	0.55+	0.73+	0.61+	ESLFLG

+ = significant at the 5% level

In table 11, we notice the highest level of agreement between the new listening test (LT1) and the first external test (WT2), and between this latter one and the new written test (WT1), followed by that between the new listening test and

Freshman English grades (FLEG). The least agreement is observed between the new written test (WT1) and ESLCE English grades (ESLEG).

A comparison of the ESLCE English grades (ESLEG) with each and combinations of the new tests shows no significant relationships, and the least agreement is observed with the new written test (WT1) and the oral test (OT1). But we see that the situation improves quite noticeably when ESLCE English grades (ESLEG) and Freshman English grades are combined.

Generally, a combination of the new tests in two's and the three in one bears the highest level of agreement with the first external test (WT2), followed by Freshman English grades (FLEG).

Given the nature and the status of the criterion measures against which the new tests are compared, the hierarchical level of agreement observed is interesting. It is interesting because it seems to suggest that the candidates' level of maturity is matched. The external test (WT2), which is a proficiency test, is for undergraduates and above. The Freshman English, which is an achievement test, is for undergraduates. And the ESLCE English grade is used both to certify high school completion and university entrance. That the new test, as proficiency test, should agree best with the external proficiency test is also encouraging.

5.4.3.4. Predictive validity

Table 12: Correlation coefficients between the new tests and the criterion grade point averages

WT1	LT1	OT1	WLT1	WOT1	LOT1	WLOT1	
0.42+	0.33	-0.12	0.35	0.15	0.30	0.23	ESLGPA
0.46+	0.03	0.36	0.30	0.28	0.22	0.32	SGPA
0.55+	0.11	0.30	0.38	0.32	0.29	0.38	ESLSGPA

+ = Significant at the 5% level

In Table 12, we notice that the new written test (WT1) has respectable and significant correlation coefficients which compare well with predictive validity mentioned by Downie & Heath (1974), Davies (1984), Kline (1986) and Hughes (1989) - 0.4. Thus, this test predicts the candidates' overall academic performance as expressed in the form of university (SGPA) and ESLCE (ESLGPA) grade point averages.

Given the evidence herein, the behaviour of the new tests is quite satisfactory; in particular, the new written test (WT1) behaved the best. The tests have acceptable reliability coefficients, and modest construct, concurrent and predictive validity. Above all, quite a reasonable proportion of respondents (teachers and students) agree that the tests are valid, especially in terms of their face and content, for assessing Freshman students' English language ability. This is, of course, very encouraging in that it supports the

attempt made to ensure relevance and representativeness of the content of the test, based on a detailed analysis of the language needs of Addis Ababa University students.

Finally, the tests have been moderated for the final administration at the end of which they will be subjected to further analysis.

CHAPTER SIX

DISCUSSION OF THE RESULTS OF THE MAIN ANALYSIS

6.1. The revised tests

The revision of the test involved mainly rejection of items and changes of format (test technique) as well as order in some limited cases. In the reading comprehension, all the true/false items and a multiple-choice question were removed. Another multiple-choice item (for two marks) is replaced by two items of identifying paragraphs that contain the given statements. While the multiple-choice items were questions on gist, the true/false items were questions on specific details of the passage. In the test of listening, the whole of 'section D' - that is, four multiple-choice items (with two options) were dropped. Unfortunately, this removal means that 'transitionals' are not tested in the revised listening test. In addition, the partial dictation is moved to the first part of the test, ie, what was 'section E' has become 'section A'. Overall, the revised test contains a total of 158 items: 124 items for the written test, 30 items for listening, and four tasks for the oral test.

6.2. Descriptive statistics

Table 13: Descriptive statistics and reliability coefficients for the revised tests (KR20)

Test	Max. Mark	No of items	Average	Stand. deviatn	coeff. of discrnm	Reliability
WT1	110	124	68.696	14.336	0.130	0.94
LT1	30	30	18.351	6.420	0.214	0.88
WT2	100	100	50.429	14.261	0.143	0.91
LT2	30	30	11.258	5.292	0.176	0.82
OT1	5	4	3.602	0.887	0.177	0.73 (KR21)

The average scores, in Table 13, show that the candidates have scored higher (done better) on the new tests which have been developed to meet their specific needs than on the international tests. In fact, the average scoring for the external test of listening is relatively low. This could be because the content of the test contains subject matter that does not have immediate relevance for the majority of our students and the fact that tape recorders were used, introducing a novel element which could have been unnerving and distracting. Either or both these reasons might account for such low achievement levels.

As expected, the average scores on all the tests turn out to be less than they were for the pilot group. In other words, the average scores for the newly designed tests have levelled off down to 63 from 70 per cent for the written test, and to 61 from 84 per cent for the listening test. There is no change in the average score of the oral test. The average score for

the external tests have also dropped to 50 from 65 per cent for the written test and to a low 38 from 43 per cent for the listening test.

The two written tests (WT1 & WT2) have very similar standard deviation, and those for the listening tests (LT1 & LT2) are not very different. While the tests are different in terms of the level of difficulty, both the newly designed and the external tests produce very similar dispersion of the candidates across the score range. Furthermore, the coefficients of discrimination show that the new listening test (LT1) spreads out the candidates more widely than the others.

6.3. Reliability

6.3.1. Test reliability

The Kuder-Richardson formulae are used to compute the reliability of the tests, ie, KR21 for the oral test and KR20 for the others. All the tests have quite satisfactory reliability coefficients (See Table 13). Note that 0.73 is considered adequate for the oral test given the brevity of the test and the conservative formula used (KR21). Moreover, the coefficients have improved when compared with the coefficients computed at the pilot stage.

6.3.2. Marker reliability

In order to have some idea of marker reliability among the language teachers involved in this study, they were required to mark and re-mark 15 paragraph scripts from the written test. These scripts were randomly picked and photocopied. The first and second sets of 15 photocopies of the paragraph scripts were given to the markers for rating at about two weeks interval. Estimates of intra- and inter-marker reliability are made based on data generated in this way. In the case of intra-rater reliability, it is estimated as the correlation between the first and second-time markings for each marker; and inter-rater reliability is also estimated as the correlation of the marks of one marker with those of another.

Table 14: Intra-marker correlation coefficients

Marker	Coefficient	Marker	coefficient
M1	0.58	M11	0.33
M2	0.63	M12	0.65
M3	0.33	M13	0.77
M4	0.53	M14	0.26
M5	0.58	M15	0.60
M6	0.63	M16	0.26
M7	0.80	M17	0.72
M8	0.69	M18	0.38
M9	0.09	M19	0.57
M10	0.62	Average	0.55

The values in Table 14 range from 0.09 for marker nine to 0.80 for marker seven, and the average coefficient is 0.55. There are only 4 (about 21%) markers (ie, 7, 8, 13, 17) with just about acceptable levels (from 0.69). The rest of the coefficients are rather low.

Table 15: Inter-marker correlation coefficients

(First one third)

	M1	M2	M3	M4	M5	M6	M7	M8	M9
2	-.09								
3	.30	.56							
4	.37	.14	.20						
5	.15	.45	.63	.20					
6	-.27	.22	.46	-.30	.49				
7	.42	.60	.71	.36	.55	.39			
8	.16	-.19	.42	.23	.16	.19	.14		
9	-.22	.35	.40	.03	.42	.33	.07	.33	
10	-.11	.51	.55	.30	.30	.51	.62	.49	.34

(Second one third)

	M1	M2	M3	M4	M5	M6	M7	M8	M9
11	.33	.40	.29	.50	.26	-.13	.25	.39	.38
12	.34	.57	.56	.33	.74	.18	.63	.40	.18
13	-.06	.03	.34	.11	.50	.44	.29	.08	.25
14	-.07	-.27	.10	-.44	.26	.39	-.08	.07	.16
15	.00	.50	.23	.29	.08	.27	.65	.17	-.18
16	.36	.12	.46	.35	.24	.07	.40	.50	.26
17	.36	.67	.76	.42	.61	.21	.78	.37	.26
18	.34	.06	.48	.36	.22	.32	.51	.46	.12
19	.32	.24	.06	.38	-.08	-.02	.41	.43	-.30

(Last one third)

	M10	M11	M12	M13	M14	M15	M16	M17	M18
11	.40								
12	.34	.39							
13	.24	.13	.27						
14	.05	.19	.04	.47					
15	.59	-.04	.25	-.12	-.39				
16	.57	.63	.34	.47	.14	.09			
17	.50	.30	.77	.06	-.30	.65	.35		
18	.54	.38	.25	.68	.26	.27	.80	.32	
19	.42	.20	.45	.02	-.24	.53	.41	.41	.50

The values range from as low as -0.44 for markers 4 and 14 to 0.80 for markers 16 and 18, with absence of correlation between marker 1 and 15. There are only six coefficients that can be considered satisfactory and the rest of the coefficients are very low.

Thus, this low level of marker reliability is rather worrying. Fortunately, and unlike the section on paragraph writing, the responses to the rest of the new written test do not require the personal judgement of the markers.

The fact that some of the ratings of the markers showed no relationship or are inversely related raises an interesting question. For example, does this situation coincide with certain aspects like training, experience or age in the markers' background information? The group is quite mixed: some had training in the teaching of literature and others in the teaching of English, some are senior and others are junior staff members as well as being both young and middle aged. Unfortunately, summarized background information on the markers is not available at this point of time to check out this question. Furthermore, the situation may be a reflection of the extent to which or the way in which markers have made use of the marking guidelines and the model (answer) paragraph. In fact, this use of or failure to use the guidelines might have affected the ratings of markers. For instance, it is likely that some markers probably did not see the need to use the guidelines when they marked the second time round assuming that they knew or remembered them. Another

possible explanation for the disappointing correlation coefficients could be the method used in gathering the data. The fact that markers probably suspected that their judgement in marking was being investigated (tested) might have in some way caused inconsistency in their ratings.

6.4. The effect of background variables

Table 16: Descriptive statistics by School, Town, Faculty, and Sex

WRITTEN TEST (WT1)						
Category	Code	Average	Stand. deviatn	Maxm.	Minm.	N
School N=202	1 Government	68.00	13.17	96.00	29.00	168
	2 Private	75.51	16.35	99.00	40.00	28
	3 Both	75.53	6.38	87.75	69.25	6
Town N=203	1 Addis Ababa	72.47	14.42	99.00	25.00	76
	2 Province	66.98	13.65	93.75	29.75	122
	3 Both	66.95	9.61	75.75	51.50	5
Faculty N=207	1 Social Science	66.06	14.84	99.00	25.00	123
	2 Science	72.55	12.70	96.00	29.00	84
Sex N=207	0 Female	72.37	12.36	93.75	51.00	16
	1 Male	68.39	14.48	99.00	25.00	191

LISTENING TEST (LT1)						
Category	Code	Average	Stand. deviatn	Maxm.	Minm.	N
School N=149	1 Government	18.06	6.41	29.00	2.00	129
	2 Private	21.75	5.70	28.00	7.00	16
	3 Both	19.00	3.46	22.00	14.00	4
Town N=148	1 Addis Ababa	22.29	4.36	29.00	8.00	56
	2 Province	16.28	6.25	28.00	2.00	89
	3 Both	16.33	8.15	22.00	7.00	3
Faculty N=151	1 Social Science	16.94	6.43	29.00	2.00	108
	2 Science	21.88	4.89	28.00	5.00	43
Sex N=151	0 Female	20.25	4.12	27.00	12.00	12
	1 Male	18.19	6.57	29.00	2.00	139

ORAL TEST (OT1)						
Category	Code	Average	Stand. deviatn	Maxm.	Minm.	N
School N=62	1 Government	3.45	0.87	5.00	1.67	45
	2 Private	3.97	0.95	5.00	2.33	13
	3 Both	4.08	0.42	4.67	3.67	4
Town N=62	1 Addis Ababa	4.15	0.78	5.00	2.33	23
	2 Province	3.23	0.77	4.67	1.67	37
	3 Both	4.33	0.47	4.67	4.00	2
Faculty N=62	1 Social Science	3.51	0.89	5.00	1.67	35
	2 Science	3.73	0.89	5.00	2.33	27
Sex N=62	0 Female	4.44	0.71	5.00	3.33	9
	1 Male	3.46	0.84	5.00	1.67	53

EXTERNAL WRITTEN TEST (WT2)						
Category	Code	Average	Stand. deviatn	Maxm.	Minm.	N
School N=166	1 Government	50.11	13.31	88.00	17.00	138
	2 Private	55.50	17.82	91.00	27.00	22
	3 Both	53.17	11.92	70.00	38.00	6
Town N=166	1 Addis Ababa	56.31	14.02	91.00	23.00	59
	2 Province	48.10	13.25	88.00	17.00	103
	3 Both	44.75	8.73	57.00	38.00	4
Faculty N=169	1 Social Science	47.82	13.01	91.00	17.00	119
	2 Science	57.24	14.57	88.00	23.00	50
Sex N=170	0 Female	49.38	13.51	70.00	28.00	16
	1 Male	50.54	14.37	91.00	17.00	154

EXTERNAL LISTENING TEST (LT2)						
Category	Code	Average	Stand. deviatn	Maxm.	Minm.	N
School N=149	1 Government	10.63	4.76	28.00	0.00	129
	2 Private	16.25	6.88	28.00	7.00	16
	3 Both	14.25	3.95	19.00	11.00	4
Town N=148	1 Addis Ababa	14.25	6.19	28.00	4.00	56
	2 Province	9.62	3.77	23.00	0.00	89
	3 Both	8.33	2.52	11.00	6.00	3
Faculty N=151	1 Social Science	10.51	4.19	28.00	4.00	108
	2 Science	13.14	7.09	28.00	0.00	43
Sex N=151	0 Female	15.92	5.63	27.00	7.00	12
	1 Male	10.86	5.09	28.00	0.00	139

Studying Table 16 we find information regarding the effect of background on the performance of candidates on the three newly designed and the two external tests.

Considering average scores, supported by information from the maximum and minimum score columns, students who come from private schools and from the city of Addis Ababa have done better on all the tests except on the oral test (OT1). Candidates with a background of both the city and the provinces, and private and government schools did better on this test. The performance of Science students is better than that of the Social Science students. Female candidates have done better than the male candidates on all the tests except on the external written test (WT2). In this latter case the male candidates are slightly in the lead.

Considering standard deviation for the spread of scores on the tests, all but the new listening test have a relatively wider spread for students who come from private schools and Addis Ababa. The listening test spreads out widely the candidates from government schools. The standard deviations of the external tests (WT2 & LT2) are relatively higher for candidates from the Faculty of Science whereas those for the local tests (WT1 & LT1) are larger for students from the College of Social Science. The spread on the oral test is the same for both faculties. The spread of scores on all the tests except the external test of listening is wider for male candidates. The spread for female candidates is wider on the external test of listening (LT2).

Although some of the differences in the average scores are slight, the analysis of variance (ANOVA) carried out to investigate the effect of these variables in the performances of candidates on these tests reveals that some of the differences are significant.

Table 17: Computed F values from analysis of variance

Test	School	Town	Faculty	Sex
WT1	1.73	0.62	9.73**	0.02
LT1	2.02	5.22**	8.23**	0.61
OT1	2.24	17.46**	1.96	11.06*
WT2	2.98	2.01	7.09*	0.14
LT2	4.14*	7.50**	3.56	9.08*

* = Significant at the 5% level

** = Significant at the 1% level

The effect of urban experience shows in the performances of candidates on the two listening tests (LT1 & LT2) and the oral test (OT1). The faculty the students come from also has an effect on their performances on the two written tests (WT1 & WT2) and the newly designed test of listening. Gender, too, has affected the performances of candidates on the oral test and the external test of listening. The effect of school shows only in the performances of candidates on the external test of listening. Generally, the effect of the variables, particularly that of town, faculty and sex, can be predicted with greater precision (ie, 1%) on the local tests. From the external tests, this greater precision is possible only on the listening test (LT2).

That urban experience should affect candidates performances on the tests that are closer to language production or language use is interesting. The higher average scores for candidates from the urban centre reflects the advantage they have gained over those from the provinces, and this may be indicative of the positive contributions of this variable to the development of language ability.

The effect of the faculty variable in the performances of the candidates on the newly designed written and listening tests, and the external written test is not easy to understand. In other words, why students from the Faculty of Science are advantaged is not very clear. In fact, there was fear that some of the source passages might favour students from the College of Social Science. What these three tests have in common can be described as 'academic English', and focusing on the local tests, both involve plenty of reading and writing. But there is very little extended writing, apart from writing a short paragraph on education, which involves classification. Broadly speaking, then, it may be that students from the Faculty of Science are more disposed to the kind of academic English sampled by the tests than are students from the college of Social Science. And this may suggest that what matters is the length of passages to be read, the type of text or writing (or rather the mode of discourse in these), and the types of tasks (activities) in the test(s) in favouring a particular category of candidates.

6.5. Pass/Fail cut-offs

In order to establish the pass/fail cut-off point, the university Semester Grade Point Average (SGPA) is thought to be the best criterion to explore from the available data at hand. The Semester Grade Point Average is computed from grades in the different subjects or courses (including English) the students have registered for for a 16-week semester. In our case, this grade point average is a grade that the candidates have earned at the end of the same semester they took the series of tests for this study. Approximately there was a two months difference between taking the tests and their respective course examinations. All university students must have a grade point average of two or better to stay on in the university.

Out of the gross 256 candidates who sat for the series of the tests, 207 (81%) of them are clear passes with a SGPA of two or above. The rest are failures, ie, 49 (19%) of them have less than two. One problem here is that the students whose grades were not available (entered) at the time of data collection are considered failures.

With this information in mind, the standard deviations and the average scores of the newly designed tests are used to decide on the cut-off points for pass/fail on these tests. In this study, the score point at the distance of the standard deviation below the average score for each test is the cut-off point for pass/fail. In other words, the average score minus

the standard deviation is the critical score point for pass/fail. For the written test (WT1), 55 (69 - 14) is the minimum score for a pass, and for the listening test (LT1), 12 (18 - 6) is the minimum score required to pass it. Accordingly, 175 (85%) candidates with a score of 55 or better pass, and 32 (15%) fail the written test. For the listening test, 125 (83%) candidates pass with a score of 12 or above, and 26 (17%) fail. Unlike these two tests, the pass/fail issue for the oral test had already been decided at the design stage when the marking guidelines and the scoring sheet were constructed. Still, 55 (89%) candidates with a score of three or better pass, and 7 (11%) fail this oral test.

According to this method, the average score remains well within the pass interval. Also the differences between the percentage of passes in the university SGPA and the new tests are not too wide. In addition, a minimum score of 55 on the written test means that the candidate has successfully answered about 44% of the items on the test. A minimum score of 12 on the listening test also means that the candidate has succeeded in answering 40% of the items on it. Thus, these minimum achievements, well over one-third of the tests, are considered satisfactory.

Besides establishing the pass/fail level, it is felt necessary to calibrate the scores on these tests according to the university letter-grade system (A,B,C,D,F). Using the same approach, the interval for each letter grade is determined based on the width of the standard deviation of each test.

More specifically, the 'C' intervals are twice as wide as the standard deviation of each test, ie, 28 for the written test and 12 for the listening. Then, the 'B' and 'D' intervals lie in the next stretch of the standard deviation above and below the average score of each test, and so on. Differently put, we first move down one stretch of the standard deviation of each test below the average score to fix the lower end, and move up the same distance from the average score of each test to fix the upper end of pass or 'C'. The 'B' interval for the written test stretches up full length and the same interval for the listening test stretches half length from the upper ends of the 'Cs'. The 'D' intervals are set in the same way down the lower ends of 'Cs'. Finally, beyond the upper limits of 'Bs' are the 'A' intervals for the really good performances of candidates, and beyond the lower limits of 'Ds' are the 'F' intervals for the clear failures on each test. Accordingly, the cut-off points for the letter grades are as follows for this group.

Written Test

A = 98 and above	(4 students)
B = 84 - 97	(30 students)
C = 55 - 83	(141 students)
D = 41 - 54	(24 students)
F = 40 and below	(8 students)
Total	207 students

Listening Test

A = 28 and above	(7 students)
B = 25 - 27	(17 students)
C = 12 - 24	(101 students)
D = 9 - 11	(10 students)
F = 8 and below	(16 students)
Total	151 students

A further investigation is to consider cut-off points for the overall performance on all the tests, ie, the written, the listening and the oral tests put together. So from the gross 256 candidates only 41 of them took all the three tests. Their scores for these were added up to examine the distribution. Like for the separate tests, the distance of the standard deviation in relation to the average score is used to fix the pass/fail limit and to set the boundaries of the intervals for the letter grades. The average score of the overall performance on the three newly designed tests is 98.76 and the standard deviation is 17.56.

Thus, the cut-off score for pass/fail is 81, ie, the average (99) minus the deviation (18). Accordingly, 36 (88%) candidates pass and 5 (12%) candidates fail the overall test. Then we move up and down the full length of the standard deviation for the other intervals. The breakdown for the group of 41 candidates is as follows.

Overall Test

A = 136 and above	(0 students)
B = 118 - 135	(5 students)
C = 81 - 117	(31 students)
D = 63 - 80	(3 students)
F = 62 and below	(2 students)
Total	41 students

Finally, the performance of each candidate can be reported specifically for each test and for the overall result. The performances of two candidates are reported below to demonstrate this. One of the candidates (*) has the highest score of 127 (97+26+4) and the other (**) has the lowest score of 47 (39+5+3) on the overall test result.

TEST	GRADE				
	A	B	C	D	F
Written Test		*			**
Listening Test		*			**
Oral Test		*	**		
Overall		*			**

6.6. Item analysis of the tests

6.6.1. Classical item analysis

The results of classical item analysis are presented in two tables for each test, that is, first in the form of a summary in terms of the components of the tests, and then for each item and section as they appear in the test papers.

Table 18: Results of item analysis for test components

WRITTEN TEST (Reliability 0.94, KR20)				
Component	Items	Facility value	Discrimn index	Item-test correln
Reading Comprehension	1-3,4-13, & 72	0.664	0.351	0.332
Discourse Indicators	14-18, 67-71	0.683	0.304	0.267
Vocabulary	19-23	0.591	0.511	0.434
Gen. Reading Comprehension	1-23, 67-71, & 72	0.658	0.362	0.328
Reference skills	24-31, 32-36	0.733	0.240	0.242
Writing	37-56, 73-78	0.796	0.304	0.280
Grammar	57-66 79-83	0.694	0.369	0.393
Cloze	84-123	0.476	0.289	0.293
Overall mean	1-123	0.640	0.314	
LISTENING TEST (Reliability 0.88, KR20)				
Partial dictation	1-15	0.710	0.550	0.515
Listening comprehension	16-23	0.508	0.613	0.515
Listening comprehension	24-30	0.519	0.376	0.330
Gen. listening comprehension	16-30	0.513	0.502	0.428
Overall mean	1-30	0.612	0.526	

The overall mean facility value for the newly designed written test (WT1) is 0.640. The facility values for the various components of the test range from 0.476 for Cloze to 0.796 for writing. The overall mean discrimination index is 0.314. The discrimination indices for the components range from 0.240 for reference skills to 0.511 for vocabulary.

The overall mean facility value for the newly designed test of listening is 0.612. The facility values for the various components of the listening test range from 0.508 for the second section to 0.710 for the partial dictation. The overall mean discrimination index is 0.526 for this test. The indices range from 0.376 for the third section to 0.613 for the second section.

The overall mean facility value for the external written test (WT2) is 0.504 and that for the external listening test (LT2) is 0.375. The discrimination index is 0.348 for the former and 0.402 for the latter.

Considering the average item-test correlations for the components or sections of the tests, we see that they contribute to the whole tests, and that there are no serious overlaps.

The overall mean facility values and discrimination indices of the newly designed tests at the final administration stand at better levels when compared with the levels at the pilot administration. That is, the overall mean facility value for the written test has dropped down to 0.64 from 0.71 and the listening test to 0.61 from 0.82. The overall mean discrimination index for the written test has risen to 0.31 from 0.21 and the listening test to 0.53 from 0.28. These desirable changes in the values of facility and discrimination suggest that these two local tests have challenged and discriminated among the candidates better at the final than

they did at the pilot stage. This behaviour compares rather encouragingly with the behaviour of the two external tests that have facility values of 0.50 and 0.38, and discrimination indices of 0.35 and 0.40 at the final administration.

The facility values of 0.64 and 0.61 for the newly designed tests compare well with the generally acceptable value of between 0.4 and 0.6, and not very far away from the most desirable value - 0.5 (Harrison, 1983; Heaton, 1975). The discrimination indices of 0.31 and 0.53 for the same tests also compare well with the satisfactory value of 0.3 or higher suggested by Baker (1989) and Harrison (1983).

Concerning the item-test correlations, all the components of the two tests meet the satisfactory level of beyond 0.2 (Kline, 1986), and most of the same components meet the satisfactory level of 0.3 or above (Hughes, 1989).

The table below is the detailed result of item analysis of the two new tests.

Table 19: Results of item analysis for each item and section

WRITTEN TEST (WT1)				
<u>Section</u>	<u>Item</u>	<u>Facility value</u>	<u>Discrimination index</u>	<u>Item-total correlation</u>
A	1	0.841	0.268	0.376
	2	0.918	0.286	0.528
	3	0.720	0.268	0.232
Average	-	<u>0.826</u>	<u>0.274</u>	<u>0.379</u>
B	4	0.696	0.339	0.283

	5	0.691	0.411	0.392
	6	0.541	0.375	0.312
	7	0.681	0.411	0.384
	8	0.599	0.482	0.370
	9	0.580	0.304	0.250
	10	0.787	0.411	0.423
	11	0.556	0.411	0.297
	12	0.633	0.393	0.340
	13	0.478	0.321	0.264
Average	-	<u>0.624</u>	<u>0.386</u>	<u>0.332</u>
C	14	0.754	0.268	0.230
	15	0.391	0.321	0.265
	16	0.734	0.00	0.017
	17	0.850	0.304	0.325
	18	0.889	0.143	0.183
Average	-	<u>0.724</u>	<u>0.207</u>	<u>0.204</u>
D	19	0.473	0.607	0.513
	20	0.609	0.661	0.503
	21	0.541	0.339	0.297
	22	0.618	0.518	0.443
	23	0.715	0.429	0.416
Average	-	<u>0.591</u>	<u>0.511</u>	<u>0.434</u>
E	24	0.652	0.304	0.235
	25	0.623	0.304	0.189
	26	0.604	0.304	0.156
	27	0.585	0.339	0.211
	28	0.899	0.250	0.348
	29	0.667	0.250	0.314
	30	0.783	0.125	0.179

	31	0.865	0.125	0.230
Average	-	<u>0.710</u>	<u>0.250</u>	<u>0.233</u>
F	32	0.531	0.321	0.260
	33	0.841	0.196	0.224
	34	0.932	0.250	0.374
	35	0.865	0.161	0.271
	36	0.681	0.196	0.148
Average	-	<u>0.770</u>	<u>0.225</u>	<u>0.255</u>
G	37	0.628	0.357	0.279
	38	0.860	0.357	0.384
	39	0.527	0.214	0.212
	40	0.889	0.250	0.302
	41	0.512	0.357	0.239
	42	0.836	0.107	0.135
	43	0.705	0.518	0.382
	44	0.850	0.339	0.300
	45	0.855	0.321	0.303
	46	0.845	0.339	0.337
	47	0.860	0.321	0.317
	48	0.758	0.464	0.387
	49	0.773	0.464	0.382
	50	0.845	0.357	0.321
	51	0.826	0.375	0.334
	52	0.836	0.375	0.335
	53	0.821	0.411	0.357
	54	0.749	0.429	0.357
	55	0.749	0.482	0.393
	56	0.647	0.536	0.399
Average	-	<u>0.769</u>	<u>0.367</u>	<u>0.323</u>

H	57	0.874	0.286	0.355
	58	0.870	0.339	0.418
	59	0.947	0.143	0.386
	60	0.522	0.500	0.377
	61	0.671	0.232	0.255
	62	0.870	0.214	0.344
	63	0.662	0.393	0.336
	64	0.802	0.429	0.380
	65	0.884	0.339	0.420
	66	0.957	0.089	0.308
Average	-	<u>0.806</u>	<u>0.296</u>	<u>0.358</u>
I	67	0.691	0.375	0.371
	68	0.681	0.446	0.363
	69	0.575	0.464	0.335
	70	0.488	0.393	0.286
	71	0.773	0.321	0.298
Average	-	<u>0.642</u>	<u>0.400</u>	<u>0.331</u>
J	72	0.580	0.232	0.202
	73	0.749	0.143	0.167
	74	0.667	0.321	0.343
	75	0.981	0.036	0.095
	76	0.986	0.036	0.189
	77	0.961	0.036	0.054
	78	0.976	0.000	-0.014
Average	-	<u>0.843</u>	<u>0.115</u>	<u>0.148</u>
L	79	0.807	0.250	0.301
	80	0.575	0.589	0.530
	81	0.357	0.589	0.455
	82	0.415	0.661	0.538

	83	0.198	0.482	0.494
Average	-	<u>0.470</u>	<u>0.514</u>	<u>0.464</u>
M	84	0.570	0.411	0.283
	85	0.778	0.304	0.325
	86	0.855	0.286	0.328
	87	0.372	0.304	0.249
	88	0.937	0.179	0.264
	89	0.850	0.321	0.404
	90	0.691	0.214	0.284
	91	0.865	0.250	0.405
	92	0.570	0.411	0.339
	93	0.126	0.125	0.189
	94	0.870	0.161	0.227
	95	0.676	0.304	0.271
	96	0.082	0.089	0.203
	97	0.300	0.054	0.096
	98	0.826	0.321	0.373
	99	0.531	0.357	0.315
	100	0.092	0.089	0.108
	101	0.290	0.429	0.323
	102	0.145	0.304	0.334
	103	0.671	0.429	0.401
	104	0.043	0.071	0.141
	105	0.599	0.375	0.353
	106	0.198	0.321	0.321
	107	0.039	0.071	0.148
	108	0.159	0.393	0.394
	109	0.758	0.071	0.126
	110	0.604	0.500	0.350

111	0.599	0.214	0.186
112	0.609	0.304	0.273
113	0.295	0.482	0.418
114	0.710	0.571	0.497
115	0.145	0.232	0.262
116	0.531	0.411	0.383
117	0.300	0.393	0.394
118	0.232	0.500	0.510
119	0.483	0.339	0.288
120	0.478	0.357	0.314
121	0.594	0.304	0.299
122	0.309	0.250	0.237
123	0.237	0.054	0.119
Average -	<u>0.476</u>	<u>0.289</u>	<u>0.293</u>
<u>Overall mean</u>	<u>0.640</u>	<u>0.314</u>	

LISTENING TEST

A	1	0.841	0.341	0.389
	2	0.596	0.561	0.461
	3	0.603	0.683	0.544
	4	0.556	0.683	0.532
	5	0.815	0.366	0.423
	6	0.821	0.390	0.466
	7	0.742	0.561	0.491
	8	0.596	0.829	0.674
	9	0.642	0.732	0.630
	10	0.848	0.415	0.519
	11	0.841	0.390	0.508
	12	0.510	0.732	0.513
	13	0.887	0.341	0.534

	14	0.649	0.488	0.446
	15	0.702	0.732	0.590
Average	-	<u>0.710</u>	<u>0.550</u>	<u>0.515</u>
B	16	0.768	0.268	0.285
	17	0.589	0.634	0.504
	18	0.437	0.732	0.573
	19	0.212	0.415	0.407
	20	0.675	0.683	0.551
	21	0.490	0.878	0.669
	22	0.662	0.780	0.672
	23	0.232	0.512	0.456
Average	-	<u>0.508</u>	<u>0.613</u>	<u>0.515</u>
C	24	0.709	0.366	0.324
	25	0.623	0.390	0.359
	26	0.662	0.171	0.203
	27	0.318	0.488	0.398
	28	0.377	0.341	0.286
	29	0.417	0.415	0.361
	30	0.530	0.463	0.377
Average	-	<u>0.519</u>	<u>0.376</u>	<u>0.330</u>
<u>Overall mean</u>		<u>0.612</u>	<u>0.526</u>	

6.6.2. Rasch analysis

Recent developments in language testing include the use of item response theory (IRT) to investigate the qualities of language tests (Adams et al., 1987; Henning, 1987; Hughes, 1989; Skehan, 1989; McNamara, 1990 & 1991). Given its advantages (Henning, 1987; McNamara, 1991) in supplying useful

and complementary information (Woods and Baker, 1985; Henning, 1987), the Rasch model has been used to assess the qualities of the tests in this study.

Among the advantages (Henning, 1987) that the Rasch one-parameter model offers are sample-free item calibration, test-free person measurement, multiple reliability estimation, and identification of guessers and other deviant respondents. From this list, multiple reliability estimation - the possibility of finding out the amount of measurement error for each item, has been the main attraction for this study. The other advantage, according to McNamara (1990 & 1991), is its potential for investigating aspects of the validity of language tests.

Henning (1987) explains IRT in the following manner:

Due to the probabilistic nature of latent trait models, it is possible to quantify for any person or item the magnitude of the departure of the given pattern of responses from the pattern predicted by the model. This departure or unlikelihood statement is a kind of response validity or model fit validity estimate that is available for both persons and items...(p.115).

In the discussion of the output from this model, McNamara (1991: 140-141) states that "Rasch IRT analysis uses data from performances of individuals on test items to construct a single measurement scale for person ability and item difficulty".

Generally, certain requirements need to be met to perform the analysis. Hughes (1989: 163) points out that "Rasch analysis begins with the assumption that items on a test have a particular difficulty attached to them, that they can be placed in order of difficulty, and that the test taker has a fixed level of ability". The application of the Rasch model requires that all items measure on the same dimension, that they are 'locally independent' and that they have equal discrimination at each decision point (Adams et al., 1987: 19-21). Thus it is said that the assumption of unidimensionality underlies Rasch model applications.

It is also pointed out that the model identifies test performance that deviates from the expected (predicted) patterns of responses (Henning, 1987; Hughes, 1989; McNamara, 1990 & 1991). That is, items or persons with values falling outside the range of +2 to -2 indicate significant deviations in terms of the model (Henning, 1987; McNamara, 1990 & 1991). It is generally expected that around 2% of candidates generate values greater than +2 (Pollitt and Hutchinson, 1987).

Below is a report of the results of the Rasch model analysis of the test data.

Table 20: Number of person and item misfits in the tests

Test	No of Person misfit > +2	%	No of Item misfit > +2	%
Written Test	36	17	44	36
Listening Test	12	8	9	30
Written Test (External)	31	18	43	43
Listening Test (External)	11	7	6	20

The results of the Rasch model analysis signal a certain number of departures from the expected patterns of responses. That is, 17% of the candidates for the written test and 8% for the listening test with misfit values greater than +2 do not conform to the model. This departure indicates that for these candidates the dimension may not be defined in the same way as for the other candidates. Again, 36% and 30% of the items in the written test and the listening test respectively are identified as misfits to the model with values greater than +2. This deviation indicates that these items are not contributing in the same way as the others to the definition of a single measurement dimension.

Comparing the new tests and the external tests, the general picture is not that different. Given that the international tests are developed for a wider range of candidates, one might expect less person and item misfits with such a wider range.

Possible causes of excessive misfit values for persons include irrelevant factors like fatigue, guessing and inattention, and surprising gaps in candidates' knowledge in the areas covered

by the test (Adams, 1987; McNamara, 1991). It may be reasonable to suspect that all or most of these factors have affected the performances of the candidates for this study. Fatigue is likely because the tests were administered during a crash-programme semester when more work in less time was required of staff and students than is usual. One evidence for this situation was the difficulty this researcher had in arranging examination times and halls (rooms). Guessing is possible, for instance, in the external tests with multiple-choice and true/false test items in the entire tests. It may also be the case that the students took the tests (or at least some of the items) less seriously because they knew that it had nothing to do with their actual academic records. The issue of surprising gaps in the candidates' knowledge of the areas tested cannot be ruled out, and it may be indicative of the existence of a heterogeneous test population.

A possible cause of excessive misfit values for items begins with poor construction (McNamara, 1991), and if items are well constructed, it has to do with item discrimination (Adams et al., 1987). It appears that weak item discrimination between low and high performers causes extreme positive values and high item discrimination causes extreme negative values (ibid.) relative to the other items. The former situation results in a number of unexpected successes or failures of candidates and the latter in a surprisingly consistent response pattern for the item. A definite explanation of the observed number of misfit items is rather difficult. But, a possible explanation that can be offered is related to the

requirements of the Rasch model. First, though, a few words about the quality of the construction of the items or the tests. A great deal of care has been taken in the construction of the items. Moreover, the tests were shown to as many language teachers as possible with one qualified tester amongst them. The tests were also tried out and piloted. As a result of these, quite a lot of changes have been made all with the intention of improving the quality of the tests. Considering the effort expended on construction, it can be claimed that the new tests are well constructed. In that case then the problem with the misfitting items could be the violation of the requirements (assumptions) of 'local independence' and 'equal discrimination' of items in the tests. A clear example that may violate the former is the labelling tasks in the tests. Each part or node to be labelled may not really be locally independent in the sense that being able to label one might help in labelling another or the rest depending on how the candidate perceives the whole task. This means every time the candidate is able to label a node he/she gains extra information useful for further response and to assess every response. This requirement is related to the assumption that all items must measure on the same dimension. Considering, for example, that the new written test is designed to test the reading and the writing skills, then it is possible that all the items are not measuring on the same dimension. McNamara's (1990) interpretation of a misfitting item supports this when he says, "...it does not form part of the same dimension as defined by other items in the test, and is therefore measuring a different construct or trait"

(p.143). In addition, it may also be the case that the items are not discriminating equally.

Concerning the newly designed tests, we see that 64% of the items of the written test and 70% of the items of the listening test are working together in contributing to the definition of a single measurement dimension for 83% of the candidates in the case of the former and for 92% of the candidates in the case of the latter.

Given that students have to be measured in the same terms as far as possible, investigation needs to continue by way of scrutinizing and interviewing misfitting items and candidates respectively. That otherwise generally satisfactory tests - on other counts such as reliability and measurement error, for instance - should yield as many misfit items and candidates as we have seen remains intriguing. With respect to the Addis Ababa University, this researcher believes that a more relaxed (generous) critical misfit level like 'three' may be a more reasonable target to achieve considering the rather restrictive circumstances we labour under.

Lastly, it is quite encouraging to see that the mean standard error of measurement is low. The mean standard error of measurement for the local written test is 0.19 ($s=.068$) and it is 0.21 ($s=.024$) for the listening test. The picture for the external tests is about the same with 0.19 ($s=.027$) for the written and 0.22 ($s=.08$) for the listening. The detailed

results of the Rasch analysis of the newly designed tests are given below.

Table 21: Rasch analysis of the written test (WT1)

Item statistics after data fit number 1

Item	Score	Difficulty	Misfit value
W1	174	-1.4590	0.9967
W2	190	-2.2614	-4.9845
W3	149	-0.6717	3.3137
W4	144	-0.5423	2.3577
W5	143	-0.5171	0.4977
W6	112	0.1952	1.8582
W7	141	-0.4673	0.3892
W8	124	-0.0685	0.7476
W9	120	0.0204	2.9194
W10	163	-1.0746	-0.6146
W11	115	0.1300	2.6171
W12	131	-0.2275	1.3265
W13	99	0.4771	2.2792
W14	156	-0.8643	2.5431
W15	81	0.8730	2.3811
W16	152	-0.7524	4.8821
W17	176	-1.5389	0.9976
W18	184	-1.9070	1.7740
W19	98	0.4988	-0.4653
W20	126	-0.1135	-0.6922
W21	112	0.1952	2.3089
W22	128	-0.1588	0.4117
W23	148	-0.6453	-0.0716

W24	135	-0.3222	1.5900
W25	129	-0.1816	1.7001
W26	125	-0.0910	2.1980
W27	121	-0.0018	1.5011
W28	186	-2.0154	-1.8336
W29	138	-0.3940	1.9776
W30	162	-1.0432	2.9802
W31	179	-1.6668	1.4469
W32	110	0.2385	1.9575
W33	174	-1.4590	2.3658
W34	193	-2.4817	-4.0087
W35	179	-1.6668	2.5902
W36	141	-0.4673	4.8945
W37	130	0.1272	1.8267
W38	178	-1.3458	-0.8614
W39	109	0.6115	2.6410
W40	184	-1.6393	-1.2772
W41	106	0.6788	2.2590
W42	173	-1.1359	4.1317
W43	146	-0.2758	-0.1105
W44	176	-1.2587	-3.0220
W45	177	-1.3017	-2.8307
W46	175	-1.2168	-3.2003
W47	178	-1.3458	-3.4915
W48	157	-0.5875	-2.3751
W49	160	-0.6807	-2.7179
W50	175	-1.2168	-3.2282
W51	171	-1.0585	-2.7595
W52	173	-1.1359	-3.3645

W53	170	-1.0210	-3.0679
W54	155	-0.5279	-1.9409
W55	155	-0.5279	-2.6170
W56	134	0.0306	-0.9788
W57	181	-1.4857	-0.6190
W58	180	-1.4377	-1.3845
W59	196	-2.5049	-3.3635
W60	108	0.6339	1.2887
W61	139	-0.0938	2.9890
W62	180	-1.4377	-0.1527
W63	137	-0.0436	1.3702
W64	166	-0.8779	0.5358
W65	183	-1.5864	-2.1172
W66	198	-2.7278	-0.5250
W67	143	-0.1966	1.5585
W68	141	-0.1448	0.8639
W69	119	0.3848	2.4516
W70	101	0.7907	3.2391
W71	160	-0.6807	2.1109
W72	120	0.3618	3.7926
W73	155	-0.5279	3.9799
W74	138	-0.0686	2.3465
W75	203	-3.5949	4.9095
W76	204	-3.8933	-2.3144
W77	199	-2.8568	6.7886
W78	202	-3.3606	13.0461
W79	166	-0.8779	-0.5021
W80	167	-0.9127	1.6965
W81	119	0.3848	0.2431

W82	74	1.4115	-0.3220
W83	86	1.1293	-0.7022
W84	41	2.3270	-2.0856
W85	118	0.0644	1.9372
W86	161	-1.0123	0.9692
W87	177	-1.5804	0.1323
W88	77	0.9641	2.0620
W89	194	-2.5644	-1.3746
W90	176	-1.5389	-1.1898
W91	144	-0.5423	2.7352
W92	178	-1.6230	0.0336
W93	118	0.0644	1.5918
W94	27	2.4560	2.6574
W95	179	-1.6668	4.3423
W96	141	-0.4673	1.8991
W97	17	2.9968	2.5334
W98	62	1.3247	4.5924
W99	170	-1.3096	0.2566
W100	110	0.2385	2.1816
W101	19	2.8702	3.9811
W102	61	1.3501	1.7258
W103	30	2.3265	-0.4992
W104	138	-0.3940	1.4282
W105	9	3.6913	2.2093
W106	124	-0.0685	1.3911
W107	41	1.9233	0.9871
W108	9	3.6913	-0.1059
W109	33	2.2068	-1.4105
W110	156	-0.8643	3.4868

W111	125	-0.0910	1.3581
W112	124	-0.0685	3.7447
W113	126	-0.1135	2.7194
W114	61	1.3501	0.4610
W115	147	-0.6192	-0.1906
W116	30	2.3265	0.2685
W117	110	0.2385	1.5535
W118	63	1.2994	0.3554
W119	47	1.7355	-1.6087
W120	101	0.4339	2.0006
W121	98	0.4988	1.7930
W122	123	-0.0462	1.9456
W123	64	1.2744	2.2903
W124	49	1.6765	3.6574

Final Rasch Item Analysis

Item	Fitted Difficulty	Standard Error	Anchored Difficulty	Scaled Difficulty
W1	-1.46	0.20	-1.46	43
W2	-2.26	0.26	-2.26	39
W3	-0.67	0.16	-0.67	47
W4	-0.54	0.16	-0.54	47
W5	-0.52	0.16	-0.52	47
W6	0.20	0.15	0.20	51
W7	-0.47	0.16	-0.47	48
W8	-0.07	0.15	-0.07	50
W9	0.02	0.15	0.02	50
W10	-1.07	0.18	-1.07	45
W11	0.13	0.15	0.13	51
W12	-0.23	0.15	-0.23	49

W13	0.48	0.15	0.48	52
W14	-0.86	0.17	-0.86	46
W15	0.87	0.15	0.87	54
W16	-0.75	0.17	-0.75	46
W17	-1.54	0.20	-1.54	42
W18	-1.91	0.23	-1.91	40
W19	0.50	0.15	0.50	52
W20	-0.11	0.15	-0.11	49
W21	0.20	0.15	0.20	51
W22	-0.16	0.15	-0.16	49
W23	-0.65	0.16	-0.65	47
W24	-0.32	0.15	-0.32	48
W25	-0.18	0.15	-0.18	49
W26	-0.09	0.15	-0.09	50
W27	-0.00	0.15	-0.00	50
W28	-2.02	0.24	-2.02	40
W29	-0.39	0.16	-0.39	48
W30	-1.04	0.18	-1.04	45
W31	-1.67	0.21	-1.67	42
W32	0.24	0.15	0.24	51
W33	-1.46	0.20	-1.46	43
W34	-2.48	0.28	-2.48	38
W35	-1.67	0.21	-1.67	42
W36	-0.47	0.16	-0.47	48
W37	0.13	0.16	0.13	51
W38	-1.35	0.21	-1.35	43
W39	0.61	0.15	0.61	53
W40	-1.64	0.23	-1.64	42
W41	0.68	0.15	0.68	53

W42	-1.14	0.20	-1.14	44
W43	-0.28	0.16	-0.28	49
W44	-1.26	0.21	-1.26	44
W45	-1.30	0.21	-1.30	43
W46	-1.22	0.20	-1.22	44
W47	-1.35	0.21	-1.35	43
W48	-0.59	0.17	-0.59	47
W49	-0.68	0.18	-0.68	47
W50	-1.22	0.20	-1.22	44
W51	-1.06	0.19	-1.06	45
W52	-1.14	0.20	-1.14	44
W53	-1.02	0.19	-1.02	45
W54	-0.53	0.17	-0.53	47
W55	-0.53	0.17	-0.53	47
W56	0.03	0.16	0.03	50
W57	-1.49	0.22	-1.49	43
W58	-1.44	0.22	-1.44	43
W59	-2.50	0.32	-2.50	37
W60	0.63	0.15	0.63	53
W61	-0.09	0.16	-0.09	50
W62	-1.44	0.22	-1.44	43
W63	-0.04	0.16	-0.04	50
W64	-0.88	0.19	-0.88	46
W65	-1.59	0.23	-1.59	42
W66	-2.73	0.35	-2.73	36
W67	-0.20	0.16	-0.20	49
W68	-0.14	0.16	-0.14	49
W69	0.38	0.15	0.38	52
W70	0.79	0.15	0.79	54

W71	-0.68	0.18	-0.68	47
W72	0.36	0.15	0.36	52
W73	-0.53	0.17	-0.53	47
W74	-0.07	0.16	-0.07	50
W75	-3.59	0.50	-3.59	32
W76	-3.89	0.58	-3.89	31
W77	-2.86	0.37	-2.86	36
W78	-3.36	0.45	-3.36	33
W79	-0.88	0.19	-0.88	46
W80	-0.91	0.19	-0.91	45
W81	0.38	0.15	0.38	52
W82	1.41	0.16	1.41	57
W83	1.13	0.15	1.13	56
W84	2.33	0.18	2.33	62
W85	0.06	0.15	0.06	50
W86	-1.01	0.18	-1.01	45
W87	-1.58	0.21	-1.58	42
W88	0.96	0.15	0.96	55
W89	-2.56	0.29	-2.56	37
W90	-1.54	0.20	-1.54	42
W91	-0.54	0.16	-0.54	47
W92	-1.62	0.21	-1.62	42
W93	0.06	0.15	0.06	50
W94	2.46	0.21	2.46	62
W95	-1.67	0.21	-1.67	42
W96	-0.47	0.16	-0.47	48
W97	3.00	0.26	3.00	65
W98	1.32	0.16	1.32	57
W99	-1.31	0.19	-1.31	43

W100	0.24	0.15	0.24	51
W101	2.87	0.24	2.87	64
W102	1.35	0.16	1.35	57
W103	2.33	0.20	2.33	62
W104	-0.39	0.16	-0.39	48
W105	3.69	0.34	3.69	68
W106	-0.07	0.15	-0.07	50
W107	1.92	0.18	1.92	60
W108	3.69	0.34	3.69	68
W109	2.21	0.20	2.21	61
W110	-0.86	0.17	-0.86	46
W111	-0.09	0.15	-0.09	50
W112	-0.07	0.15	-0.07	50
W113	-0.11	0.15	-0.11	49
W114	1.35	0.16	1.35	57
W115	-0.62	0.16	-0.62	47
W116	2.33	0.20	2.33	62
W117	0.24	0.15	0.24	51
W118	1.30	0.16	1.30	56
W119	1.74	0.17	1.74	59
W120	0.43	0.15	0.43	52
W121	0.50	0.15	0.50	52
W122	-0.05	0.15	-0.05	50
W123	1.27	0.16	1.27	56
W124	1.68	0.17	1.68	58
Average		<u>0.19</u>		

Table 22: Rasch analysis of the listening test (LT1)

Item statistics after data fit number 1

Item	Score	Difficulty	Misfit value
L1	127	-1.4553	4.4439
L2	90	0.1578	0.7312
L3	91	0.1226	1.2167
L4	84	0.3682	1.2866
L5	123	-1.2277	1.1356
L6	124	-1.2824	1.2341
L7	112	-0.6963	-0.3386
L8	90	0.1578	-2.6421
L9	97	-0.0954	-1.8434
L10	128	-1.5163	-2.6644
L11	127	-1.4553	0.7983
L12	77	0.6081	0.2083
L13	134	-1.9301	-4.6749
L14	98	-0.1326	1.3340
L15	106	-0.4431	-0.8678
L16	116	-0.8770	3.9139
L17	89	0.1929	0.3258
L18	66	0.9829	-1.1085
L19	32	2.3011	-1.1588
L20	102	-0.2848	-1.1727
L21	74	0.7102	-2.5627
L22	100	-0.2081	-2.8125
L23	35	2.1621	-1.5069
L24	107	-0.4838	3.7283
L25	94	0.0145	2.8208
L26	100	-0.2081	5.2253
L27	48	1.6277	2.6210
L28	57	1.2976	4.1223

L29	63	1.0872	3.6230
L30	80	0.5057	2.1915

Final Rasch Item Analysis

Item	Fitted Difficulty	Standard Error	Anchored Difficulty	Scaled Difficulty
L1	-1.46	0.25	-1.46	43
L2	0.16	0.19	0.16	51
L3	0.12	0.19	0.12	51
L4	0.37	0.19	0.37	52
L5	-1.23	0.23	-1.23	44
L6	-1.28	0.24	-1.28	44
L7	-0.70	0.21	-0.70	47
L8	0.16	0.19	0.16	51
L9	-0.10	0.20	-0.10	50
L10	-1.52	0.25	-1.52	42
L11	-1.46	0.25	-1.46	43
L12	0.61	0.19	0.61	53
L13	-1.93	0.28	-1.93	40
L14	-0.13	0.20	-0.13	49
L15	-0.44	0.20	-0.44	48
L16	-0.88	0.22	-0.88	46
L17	0.19	0.19	0.19	51
L18	0.98	0.19	0.98	55
L19	2.30	0.22	2.30	62
L20	-0.28	0.20	-0.28	49
L21	0.71	0.19	0.71	54
L22	-0.21	0.20	-0.21	49
L23	2.16	0.21	2.16	61
L24	-0.48	0.20	-0.48	48

L25	0.01	0.19	0.01	50
L26	-0.21	0.20	-0.21	49
L27	1.63	0.20	1.63	58
L28	1.30	0.19	1.30	56
L29	1.09	0.19	1.09	55
L30	0.51	0.19	0.51	53
Average		<u>0.21</u>		

6.7. Test validity

6.7.1. Face validity

Candidates and language teachers were required to give their views on the three newly designed tests concerning the general quality of the tests as a whole. The questionnaire items (Appendix 16-19) in this section focus on whether the items appear to be valid as well as on other issues considered important. The report contains the number of responses in two of the cells, ie, the counts for 'agree' and 'strongly agree' put together. Note again, any value greater than 33% is considered satisfactory. The contents of the questionnaire items are the same for the different tests and categories of respondents. But the difference that occurs in some limited cases is signalled by the use of dots, slashes or parentheses in the statements when reporting.

Broadly speaking, a substantial proportion of the students and teachers involved in this study have expressed positive views about the face validity of the tests.

The following is a summary of the questionnaire statements and the views of candidates and language teachers in actual counts and percentages.

Table 23: Summary of views of candidates and teachers

(Teachers: N=19; Students: N=207 [WT1], N=151 [LT1], N=62 [OT1])

Statement	WT1 CNT	%	LT1 CNT	%	OT1 CNT	%
Writing this test is enjoyable.						
1,1,1. students	163	79	111	74	51	82
This test assesses or measures...language ability/...ability to listen and understand spoken language/ ability to speak in English.						
2,2,2. students	166	80	121	80	46	74
1,1,1. teachers	17	90	15	79	13	68
This test has caused me fear, anxiety or disappointment.						
3,9,9. students	160	77	95	63	44	71
This test is (positively) different from other English language examinations like the ESLCE and/or Freshman exams.						
4. students	157	76	NA		NA	
2. teachers	18	95	NA		NA	

Table 24: Summary of students' opinion on the components of
the tests

The written test (N=207)

Statement	Count	%
1. The reading passage, pp.1-2.	157	76
2. Section A (items 1-3).	162	78
3. Section B (items 4-7).	145	70
4. Section C (items 8-12).	154	74
5. Section D (items 13-17).	151	73
6. Section E (the library catalogue cards, p.5).	140	68
7. Section E (items 18-22).	154	74
8. The dictionary page, p.6.	159	77
9. Section F (items 23-27).	151	73
10. Section G (Your role, p.7).	131	63
11. Section G (the course offering, p.8).	116	56
12. Section G (completing registration form, p.9).	124	60
13. Section H (items 28-37).	160	77
14. Section I (items 38-42).	164	79
15. Section J (the paragraph, p.12).	164	79
16. Section J (items 43-44).	162	78
17. Section K (the diagram, p.13).	156	75
18. Section K (item 45).	158	76
19. Section L (item 46).	153	74
20. Section M (items 47-86).	131	63

The listening test (N=151)

1. Section A (the partial dictation).	94	62
2. The listening passage read out.	85	56
3. Section B (items 1-2).	79	52
4. Section C (items 3-6).	88	58

The oral test (N=62)

1. Section A (the dialogue you read out with your examiner).	51	82
2. Section B (the double dialogue where you choose and respond to your examiner).	45	73
3. Section C (the class schedule).	40	65
4. Section C (the asking and answering of questions).	49	79

Notice that the lowest percentages are for the listening test, and it should be interesting to find out in a future study why the candidates did not like the test.

6.7.2. Content validity

The content of the test is derived from an extensive needs analysis. However, in addition to this and the expert opinion of one tester, the views of language teachers are sought to establish the quality of the tests in this respect.

In the third part of the questionnaire for teachers (Appendix 19), each section (with item numbers) of the tests is itemized with a statement

of what it is intended to test. The teachers are then required to indicate the extent to which they agree or disagree with the statements on a 5-point scale. Thus, information gained from this section of the questionnaire shows that a substantial proportion of language teachers agree that the tests have good content validity. The details of the positive views of language teachers in actual counts and percentages against the statements are presented below.

Table 25: Summary of teachers' views

The written test

Statement	Count	%
1. Section A (items 1-3) and section J (item 43) test skimming skill and the ability to distinguish main ideas.	18	95
2. Section B (items 4-7) tests scanning skill.	16	84
3. Section C (items 8-12) tests the recognition of indicators in discourse.	14	74
4. Section D (items 13-17) tests the ability to deduce meanings of unfamiliar words from the context.	17	90
5. Section E (items 18-22) and section F (items 23-27) test reference skills, ie, understanding and using reference sources.	17	90
6. Section G (completing a registration form) tests the ability to understand and use information/instruction to carry out a task.	17	90

7. Section H (items 28-37) and section L (item 46) test structural accuracy.	16	84
8. Section I (items 38-42) tests the ability to use indicators in discourse.	15	79
9. Section J (item 44) and section K (item 45) test the ability to transcod information, ie, from diagrams/graphs/ tables to writing and vice versa.	19	100
10. Section M (items 47-86) tests integrated reading comprehension.	17	90

The listening test

1. Section A (the partial dictation) tests the ability to discriminate sounds and write it down accurately.	16	84
2. Section B (items 1-2) tests the ability to distinguish the main idea and the ability to listen for specific details.	15	79
3. Section C (items 3-6) tests the understanding of conceptual meanings.	14	74

The oral test

1. Section A (the dialogue) tests the ability to speak (read aloud) with ease and clarity.	14	74
2. Section B (the double dialogue) tests the ability to respond appropriately with ease and clarity.	11	58

3. Section C (asking and answering questions from sources) tests the ability to ask and answer questions with ease and clarity. 12 63

In Table 25, we see that the lowest percentages are for the oral test. This again suggests that it would be useful in a future study to find out why some teachers had reservations about this test.

6.7.3. Construct validity

Table 26: Intercorrelation coefficients between tests

(WT1: N=207; LT1: N=151; OT1: N=62; WT2: N=170; LT2: N=151)

	WT1	LT1	OT1	WT2
WT1	-	-	-	-
LT1	0.72*	-	-	-
OT1	0.52*	0.48*	-	-
WT2	0.75*	0.70*	0.47*	-
LT2	0.42*	0.44*	0.57*	0.54*

* = Significant at the 5% level.

An overview of the intercorrelation coefficients for all the tests shows that there is a significant relationship between the tests, and the coefficients range from 0.42 for the new written test and the external listening test to 0.75 for the same test and the external written test. This represents a variance overlap ranging from 18% to 56%. Although some degree of integration is allowed, the new tests are intended to

measure different areas and skills in the English language. This intended difference emerges clearly in some cases, but only marginally in others. The fact that the written test (WT1) correlates highly with the external written test (WT2) than it does with the external listening test, that this last test (LT2) correlates (slightly) better with the new listening test (LT1) than it does with the new written test, and that the oral test (OT1) correlates better with the external listening test (LT2) than it does with the external written test are all pieces of evidence for the construct validity of the newly designed tests.

Focusing on the relationship between the newly designed tests, we find that the highest correlation is between the written and the listening tests (0.72), followed by the written and the oral tests (0.52); and the lowest correlation is between the listening and the oral tests (0.48). This picture reflects the extent or amount of integration allowed at the design stage. For example, if we consider the amount of reading carried out during the process of responding to items in relation to the size (length) of the tests the order (ie, from the highest to the lowest) is the written test, followed by the listening test, and this followed by the oral.

The general picture gained from the intercorrelations of the local tests clearly suggests that the tests have enough in common to be significantly related but not so much as to render one another redundant. The coefficients show that the tests measure different areas of language ability, and this

supports or confirms the realisation of the intention at the design stage. It can also be seen that prediction between the tests is possible to some extent as the lowest coefficient (0.48) between the listening and the oral tests is better than the usually mentioned level of 0.4 for this purpose (Downie and Heath, 1974; Davies, 1984; Kline, 1986; Hughes, 1989).

One piece of evidence stands out in the comparison between the local and the international tests. This is the significant and meaningful overlap between the new written test and the external written test (0.75). The external written test covers structure, written expression, vocabulary and reading comprehension, and the new written test covers these and more - like reference skills, writing a paragraph and short answers etc. Since the new test aims to test more than the external test it would actually be disappointing if it correlated very highly with the external test. In the light of this, the substantial relationship (56%) is reasonably satisfactory. Given this relationship, we can safely say that the new written test measures what it is designed for - reading and writing. The low correlation between the new written test and the external listening test (0.42) strengthens this evidence.

The picture for the new listening test is not so straightforward. It correlates more highly with the other tests than it does with the external test of listening. This is not surprising in the light of the difference observed in the design of these two listening tests. For instance, one clear source of difference between the two tests is the inclusion of

partial dictation in the new listening test. It is reassuring, however, that the low level of correlation is significant and that the coefficient between the listening tests (0.44) is slightly higher than that between the new written test (WT1) and the external listening test (0.42). Furthermore, the fact that the coefficient between the listening tests (0.44) is lower than that between the two external tests (0.54) suggests the existence of a listening bias in the new test (LT1).

There is a relatively balanced situation when it comes to the oral test. It has not more than 33% (53% when corrected for attenuation) shared variance with the other tests.

Generally, the strength of the relationship among the three newly designed tests suggests that each test contributes something to what they all measure in common as well as something specific each test is measuring. This picture of the contribution of the tests is welcome in the light of the overall assessment of the candidates.

The intercorrelations between the various sections of the three newly designed tests are presented in Table 27 to provide further information on the construct validity of the tests.

Table 27: Intercorrelation coefficients between the sections of the new tests

	RC	DI	VOC	RS	WRI	GRM	CLZ	PD	LC	OSA	OSB
Reading Comprehension	-	-	-	-	-	-	-	-	-	-	-
Discourse Indicators	0.27*	-	-	-	-	-	-	-	-	-	-
Vocabulary	0.32*	0.50*	-	-	-	-	-	-	-	-	-
Reference Skills	0.31*	0.22*	0.30*	-	-	-	-	-	-	-	-
Writing	0.52*	0.34*	0.33*	0.40*	-	-	-	-	-	-	-
Grammar	0.44*	0.50*	0.48*	0.40*	0.56*	-	-	-	-	-	-
Cloze	0.52*	0.47*	0.49*	0.42*	0.50*	0.64*	-	-	-	-	-
Partial Dictation	0.46*	0.50*	0.43*	0.35*	0.50*	0.56*	0.66*	-	-	-	-
Listening Comprehension	0.38*	0.30*	0.30*	0.30*	0.40*	0.50*	0.57*	0.64*	-	-	-
Oral Section A	0.20	0.23	0.35	0.15	0.32*	0.32*	0.38*	0.40*	0.36*	-	-
Oral Section B	0.17	0.32*	0.27	0.27	0.23	0.42*	0.25	0.32*	0.28	0.46*	-
Oral Section C	0.20	0.30*	0.50*	.07	0.28*	0.39*	0.43*	0.38*	0.39*	0.49*	0.36*

* = Significant at the 5% level.

The correlation coefficients range from as low as 0.07 between the third section of the oral test and reference skills to 0.66 between cloze and partial dictation - an overlap of about 44%. In fact, the more noticeable coefficients (for example, beyond 0.5) involve one of these latter sections in the written and listening tests respectively, except the 0.56 coefficient for writing and grammar. Most of these correlation coefficients are acceptable according to Ingram (1977) and Wall et al. (1994) who suggest that correlations between subtests should only be moderate, and not very high. The overall picture is that while the various sections are significantly related to each other they are also measuring different areas of language and language skills.

Correction for attenuation was carried out to see if any thing different might emerge in terms of the relationship between the tests. As can be seen in Table 28 below, the general picture remains the same although the coefficients have risen. The strength of the association between the two written tests (ie, WT1 & WT2) has come out more clearly.

Table 28: Intercorrelation coefficients between tests:
Corrected for attenuation

	WT1	LT1	OT1	WT2
WT1	-	-	-	-
LT1	0.79	-	-	-
OT1	0.63	0.60	-	-
WT2	0.81	0.79	0.58	-
LT2	0.48	0.52	0.73	0.62

6.7.4. Concurrent validity

Table 29: Intercorrelation coefficients between tests and grades

	WT1	LT1	OT1
WT2	0.75*	0.70*	0.47*
LT2	0.42*	0.44*	0.57*
SLEG	0.40*	0.46*	0.50*
101	0.68*	0.63*	0.47*

* = Significant at the 5% level.

An overview of the correlation coefficients in Table 29 shows that the greatest agreement is between the written tests (WT1 & WT2), followed by that between the external written test and the new test of listening. Note that the coefficients of the local and external tests when corrected for attenuation are higher than they are in this table (See Table 28). Furthermore, the newly designed written and listening tests agree better with the Freshman English examination, ie, with an exam administered at the level for which these new tests are designed, than they do with the Ethiopian School Leaving English grade.

It is encouraging to see that the three new tests agree better with the external tests which are developed with more or less the same purpose. In other words, the order of agreement observed, ie, first with the external tests, then with the Freshman English examination and lastly with the Ethiopian school leaving English grade suggests the relevance or

Statement	WT1		LT1		OT1	
	CNT	%	CNT	%	CNT	%
The instructions/directions in this test are clear.						
5,3,3. students	197	95	110	73	52	84
3,2,2. teachers	17	90	16	84	14	74
The markability of this test is satisfactory.						
5,4,4. teachers	16	84	13	68	11	58
The format or the lay out (the arrangements & conditions) of this test is (are) good and convenient.						
6,4,4. students	158	76	114	76	46	74
4,3,3. teachers	17	90	15	79	11	58
The time allotted for this test is fair or just about right.						
7,5,5. students	173	84	90	60	51	82
6,5,5. teachers	14	74	13	68	13	68
The length of this test is reasonable or just about right.						
8,6,6. students	132	64	101	67	47	76
7,6,6. teachers	9	47	15	79	12	63

Statement	WT1		LT1		OT1	
	CNT	%	CNT	%	CNT	%
The passages and sources of information in this test are interesting.						
9,7,7. students	139	67	101	67	51	82
8,7,7. teachers	14	74	13	68	15	79
The passages and the sources of information in this test are simple enough for Freshman students.						
10,8,8. students	120	58	77	51	42	68
9,8,8. teachers	18	95	13	68	15	79
The grade I will get (the candidate gets) from this test will be fair in that it should closely reflect my (his/her) language ability.						
11,10,10. students	118	57	94	62	39	63
10,9,9. teachers	14	74	11	58	11	58
Overall, this test is suitable for Freshman students.						
11,10,10. teachers	17	90	14	74	15	79

In Table 23, fewer than 60% of students agree that the passages and the sources of information in the written test as well as the listening test are simple enough for Freshman students, and that the grade they get from the written test is

a fair reflection of their language ability. It may be that the former points to the problem of striking a balance between the level of difficulty and interest in the selection of passages, and the latter to the notion of 'fairness'. Some candidates probably (and rightly) felt that in addition to the validity of the test the need for 'reliable marking' is lacking in the statement. Again, fewer than 60% of language teachers agree that the markability of the oral test is satisfactory, and that the arrangements and conditions of this same test are good and convenient.

These lower levels of agreements are clear pointers to the difficulties involved in the issues, and it would be useful to investigate into these areas in a future study.

Moreover, the opinion of students is also sought regarding the acceptability of each section or component of the tests. In this section of the questionnaire for students (Appendix 16-18), the various sections (with item numbers) and parts of the tests (like passages and sources of information) are itemized for students to judge as 'good' or 'bad' on a 5-point scale. Considering actual counts of the cells for 'good' and 'very good', a substantial proportion of students think all the sections and parts are good. The percentages range from 52% for the labelling task in the listening test to 82% for the first dialogue in the oral test. The questionnaire statements and the opinions of the students are presented below.

appropriateness of the new tests for the educational level for which they are designed.

6.7.5. Predictive validity

Table 30: Intercorrelation coefficients between the new tests and grade point averages

	WT1	LT1	OT1
SLGPA	0.33*	0.41*	0.21*
SGPA	0.60*	0.60*	0.43*

* = Significant at the 5% level.

The correlation between the three newly designed tests and the two grade point averages are considered for purposes of prediction. Both the written test and the listening test correlate with the University Semester Grade Point Average (SGPA) quite satisfactorily at $r=0.6$, the best coefficient that can be hoped for according to Downie and Heath (1974), for example. Moreover, the oral test also correlates reasonably well with the same criterion ($r=0.43$). The listening test also correlate satisfactorily with the Ethiopian School Leaving Grade Point Average (SLGPA).

Intercorrelations between the various sections of the tests and the whole tests are examined and presented below. According to Table 31, the cloze (0.89) and the grammar (0.81) sections of the written test correlate more highly with the test of reading and writing than the other sections. Furthermore, the sections of the listening test, namely, partial dictation (0.91) and listening comprehension (0.90)

also correlate very highly with the whole test. Section B (0.86), followed by section A (0.75) of the oral test also correlate highly with the total test.

Table 31: Intercorrelation coefficients between sections and whole tests

	Written Test	Listening test	Oral test
Reading Comprehension	0.70*	0.47*	0.24
Discourse Indicators	0.62*	0.39*	0.37*
Vocabulary	0.63*	0.41*	0.45*
Reference Skills	0.56*	0.36*	0.23
Writing	0.72*	0.50*	0.34*
Grammar	0.81*	0.59*	0.49*
Cloze	0.89*	0.68*	0.42*
Partial Dictation	0.72*	0.91*	0.45*
Listening Comprehension	0.59*	0.90*	0.41*
Oral Sec. A	0.41*	0.42*	0.75*
Oral Sec. B	0.38*	0.33*	0.86*
Oral Sec. C	0.47*	0.42*	0.72*

Again, Ingram (1977) writes that correlations between subtests and total test scores should be high. While most of the coefficients are acceptable according to this requirement, those for discourse indicators, vocabulary, and reference skills are doubtful. But even these sections correlate better with the whole test than with any of the other sections.

CHAPTER SEVEN

CONCLUSIONS AND RECOMMENDATIONS

7.1. Conclusions

This study has explored the state of affairs in the testing of English and the feasibility of a different approach to the testing of this language in the context of the Addis Ababa University, Ethiopia.

Initial explorations indicated the need for improvement both in the quality of the tests and the testing scheme in order to meet the demands of the situation. Thus, the study was aimed at designing, developing and validating an appropriate test to suit the circumstances of this university. The results of the study confirm the suitability of the (series of) tests which are described as tests of English for tertiary education.

Investigation into six years (1985/86 - 1990/91) used Freshman English examinations revealed that the aim of the examinations boils down to the assessment of general English, in which the assessment of structural accuracy takes the lion's share of the total marks for each examination. The test items, either in form or content, are not related to the academic activities students experience during their stay in the university. In fact, it was found that the examinations involve mostly objective-type exercise/items like true or false, multiple-choice and its variants. Thus, it is questionable whether the

examinations tested a variety of relevant skills and whether there was a stable relationship between the productive-receptive skills. In addition, some of the components of the examinations encouraged rote memorization as a method of preparation for the examinations. The two Freshman English examinations (101A & 101B) did not appear to be testing anything different between them; and this suggests the need to reconsider the aims and scopes of the two Freshman English courses.

There has not been any indication of pretesting or any occasional test investigation carried out to date. Thus, there is no information on the qualities of the items, or on issues like test and marker reliability, etc. Also the grades on these examinations hardly tell us anything apart from the relative performance of the students in relation to the others.

The contents, the aims and objectives of the newly designed tests are derived from extensive needs analyses; and therefore, they are goal-oriented, skills- and task-based. Hence, these tests take into account the communicative language needs of the students at the Addis Ababa University. The grades on these tests give an indication of the ability of the students to cope with the language demands they are likely to face or experience in their academic endeavour. For this reason, the tests and the grades on these tests are meaningful and informative both to the students and the institution. Furthermore, the relative weakness or strength of students who

have taken these tests can be identified in terms of the major language skills.

The reliability of the tests is quite satisfactory: the written test has a reliability coefficient of 0.94, the listening test has 0.88 and the oral test has 0.73. However, the intra- and inter-marker reliability study for the paragraph writing section of the written test did not show a satisfactory correlation in most of the cases.

The results of item analysis are satisfactory: the level of difficulty and discrimination of the sections and the tests are within acceptable ranges. The average facility value of the written test is 0.640 and 0.612 for the listening test; and the average discrimination index for the former is 0.314 and 0.526 for the latter. On the whole, both the various sections and the tests are all of a reasonably suitable level of difficulty and discrimination. In addition, almost all the item-test correlation indices are satisfactory; and the average correlations for the various sections range from 0.242 to 0.434 for the written test and from 0.330 to 0.515 for the listening test. This is welcome as it indicates that each section is contributing something to the total test but that no one section could replace it.

The Rasch model analysis has identified a number of departures in terms of misfitting items and candidates. Accordingly, 64% of the items of the written test and 70% of the same in the listening test are working together in contributing to the

definition of a single measurement dimension for 83% and 92% of the candidates respectively. In this study, the appropriateness of the application of the Rasch analysis to this data is investigated.

The validity of the tests is satisfactory on the whole. The face and content validities of the tests are quite satisfactory: a substantial proportion of students and language teachers involved in this study have expressed positive views about the face validity of the tests, and a similar proportion of language teachers agree that the tests have good content validity.

The intercorrelations between the various sections of the tests provide evidence for the construct validity of the tests. Since the different sections are intended to measure different aspects of language and/or language skills, they are not expected to correlate very highly with one another (Ingram, 1977; Wall et al., 1994). Scrutinized in this sense, the intercorrelation coefficients for the sections are satisfactory. That is evidence of the fact that they all measure the English language in common but also different aspects or skills of the language. The correlation coefficients between the tests also show evidence in support of the construct and concurrent validities of the tests; and this evidence is particularly clear for the written test and the oral test. In addition, the predictive validity of the tests is satisfactory; this is particularly so for the written test and the listening test.

Given the correlations between the sections and the tests, some of the values suggest the possibility of putting together a shorter test for each of the tests. For example, a shorter version of the written test could be composed of the cloze, the grammar and the writing sections, all of which have better than 50% overlaps with the actual written test. Moreover, given the amount of shared variance between the written test and the other two tests - 0.79 with listening and 0.63 with the oral test, it is possible to compose a shorter version of the written test that is reasonably likely to predict performance in the other tests. The level of agreement is encouraging for further experiment on such a prospect.

Such a short version of the test could then be used as an entrance and/or a placement test. It could be administered at entry to the university, and the performance of the candidates could be used for decisions regarding admissions and exemptions. Those candidates who really do well in the test could be exempted from taking the first part or both of the Freshman English courses and (be) placed in the Sophomore English course.

This is attractive considering the fact that the test is based on the communicative language needs of Addis Ababa University students. In other words, decisions on admissions and exemptions will have been made based on a measure of the language facility (language proficiency) the candidates have for the language demands they are to face as students in the

university, and not on the basis of achievement or performance on a previous syllabus or course of instruction in English. Broadly speaking, the qualities of the new tests are quite satisfactory. They are reliable and valid, and therefore suitable for the assessment of Freshman students' performance in English.

This study has been concerned with the development and validation of a suitable test of English for Addis Ababa university students. According to Messick (1988: 13), " ...validity is an evolving property and validation is a continuing process." In this sense, what has been done and presented herein can only be considered an interim validation of the test. With further administration(s) of the test, or if and when the test is used with the population it is intended for, more evidence is likely to emerge in support (or otherwise) of what has been established at this stage.

Finally, this study has contributed to a validation of a battery of tests of English for academic purposes to assess the English language disposition of Freshman students at Addis Ababa University, Ethiopia.

7.2. Recommendations

Testing committee: The Freshman English Committee needs to be reorganised and strengthened for the additional responsibility of carrying out investigations into the qualities of tests.

Pretesting: In the tradition of writing fresh tests or examinations every semester and every year, the tests need to be tried out on small samples before they are administered to the actual target population. The analysis of the test data will throw light on the behaviour of the tests, and the necessary improvements can then be made based on this information.

Test analysis: Preferably in addition to pretesting, actual samples of student scripts need to be analyzed to learn about the quality of the test(s) after administration. Information gained from this could be useful for test equivalence.

Test equivalence: Again, in the tradition of writing fresh tests every semester and writing different versions for different categories of students, for instance regular and extension, the issue of test equivalence is relevant and important.

Item banking: The use of this procedure of collecting test items which have been administered and analyzed is advocated here given the population of first year students taking the Freshman English examination every semester. Availability of

computer hardware and software for appropriate storage and retrieval of items is required for this procedure. Further, the Rasch One-parameter Logistic Model is advocated as an appropriate measurement model for the computation of item difficulty estimates and associated standard errors of measurement (see Henning, 1987).

Training/awareness raising: Both the department and the individual teachers need to find ways and means of improving on the skills of writing tests, and more importantly on marking with consistency. Talks, seminars, workshops and discussions should be organised more frequently. Arrangements should be made for marking samples of scripts of written language and rating samples of video (tape) recordings of spoken language in groups and pairs. Innovative, fresh or novel ideas should be considered and encouraged in these activities. Individual teachers would also do well if they tried to check their own marker/rater reliability on some regular basis. It is advisable that teachers form small groups with fellow teachers for purposes of checking and maintaining intra- and inter-marker reliability. It is suggested that this be done using material from actual tests for actual evaluation.

Testing listening and speaking: This study has demonstrated the need to test these skills and explored some ways of testing them. One of the usual problems with the testing of the listening skill is lack of tape recorders, and this can be avoided, as in this study, using live human voice. Although

some effort has been put into the testing of speaking in this study, the time it takes to assess oral production remains a problem. Several options should be explored further until a fairly satisfactory solution is found. For the moment, it seems appropriate to assess these skills in classroom situations. The Freshman English Testing Committee should find ways of providing assistance while acting as a moderator.

Testing speaking: The first important thing that needs to be done, at the level of the department and language teachers, is to carefully reconsider the issue and decide whether testing this skill is worthwhile considering the excessive time and arrangements it demands. Where a skills-based test of English, skills-based course objectives, and an overall evaluation of students is desirable - as advocated in this study - the position of this researcher is in favour of testing the speaking skills of Freshman students. Its inclusion in the overall evaluation of students could improve the overall validity of assessment, and produce some general motivational effect (Walker, 1990).

Some specific options that could be explored include :

First, assessment of the oral skill could take place during scheduled teaching hours. The teacher may have to use the last few teaching hours of each section (group) for the purpose of oral assessment. eg. 40 students multiplied by 10 minutes of test = 400 minutes of oral testing. This takes up 8 hours of teaching, ie, 400 divided by 50 minutes of teaching hour.

Second, since the rating of oral production is better done by more than one person, arrangements could be made to group raters and conduct the test over a period of one week. The last teaching week could become an official oral-test week. This is partly inspired by a research report on large-scale oral testing (Walker, 1990). In this experimental research, there were 20 testing staff for a population of 1000 students taking an eight-minute oral test. It took them two days (16 hours). Given that our student population is at least twice as much, we will need approximately double this time.

Lastly, this researcher believes that the tutorial hours could provide us with the flexibility of time and space we would need for the testing of the listening and speaking skills.

Entrance examination: This researcher is aware of the fact that university entrance examinations have been considered for some time now. It is believed here that, in the case of English (entrance) examination, the idea of exemption or placement should also be considered as well.

BIBLIOGRAPHY

- Adams, R.J. et al. 1987 "A latent trait method for measuring a dimension in second language proficiency," Language Testing, 4/1, 9-27.
- Alderson, J.C. 1981. "Report of the discussion on communicative Language testing," in Alderson, J.C. and Hughes, A. (eds.) pp. 55-65.
- _____ 1988. "Testing English for specific purposes: How specific can we get?" in Hughes, A. (ed.) pp. 16 - 28.
- Alderson, J.C. et al. 1987. "An Evaluation of the National Certificate in English: Sri Lanka 1986," Final Report, Vol. 2. Lancaster: Institute for English Language Education, University of Lancaster.
- Alderson, J.C. and Hughes, A. eds. 1981. ELT Documents 111...Issues in Language Testing. London: The British Council.
- Alderson, J.C. and Urquhart, A.H. 1984. "ESP tests: The problem of student background discipline," Occasional papers 29: Practice and Problems in Language Testing. Culhane, T. et al. (eds.) University of Essex, pp. 1-13.
- Bachman, L.F. 1990. Fundamental Considerations in Language Testing. Oxford: Oxford University Press.
- Bachman, L.F. 1982. "The Trait Structure of Cloze Test Scores," in Oller, J.W. and Jonz, J. (eds.) pp. 177-188.

- Bachman, L.F. et al. 1988. "Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries," Language Testing, 5/2, 128-159.
- Baker, D. 1989. Language Testing. London: Edward Arnold.
- Brown, J.D. 1983. "A Closer Look at Cloze Validity," in Oller, J.W. and Jonz, J. (eds.) pp. 189-196.
- Canale, M. 1984. "On some theoretical frameworks for language proficiency," Language Proficiency and Academic Achievement. Rivera, C. (ed.) Clevedon: Multilingual Matter Ltd, pp. 28-40.
- Conroy, K. ed. 1985. Freshman English 101A. Addis Ababa: Institute of Language Studies, Addis Ababa University.
- _____ ed. 1986. Freshman English 101B. Addis Ababa: Institute of Language Studies, Addis Ababa University.
- Criper, C. and Davies, A. 1988. ELTS Validation Project Report(ii). The British Council.
- Cummins, J. 1984. "Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students," Language Proficiency and Academic Achievement. Rivers, C. (ed.) Clevedon: Multilingual Matters Ltd, pp. 2-19.
- Curry, D. 1988. More Dialogue for Everyday Use. Washington, D.C.: English Language Programs Division, United States Information Agency.

- Davies, A. 1988. "Communicative Language Testing," in Hughes A. (ed.) pp. 5-15.
- _____ ed. 1968. Language Testing Symposium: a psycholinguistic approach. London: Oxford University Press.
- _____ 1990. Principles of Language Testing. Oxford: Basil Blackwell.
- _____ 1984. "Validating three tests of English language proficiency," Language Testing, 1/1, 50-69.
- Dejene Leta. 1990. "Achievement, Washback, and Proficiency In School Leaving Examination: A case of Innovation In An Ethiopian Setting," Ph.D. Thesis. Lancaster: University of Lancaster, Department of Linguistics and Modern English Language. (Unpublished)
- Downie, N.M. and Heath, R.W. 1974. Basic Statistical Methods, Fourth edition. New York: Harper & Row, Publishers.
- Eshetie, A. 1974. "The Pre-war Attempts to Promote the Use of the English Language in the Educational System of Ethiopia in Place of French," The Ethiopian Journal of Education, VI, 65-84.
- Feare, R.E. 1989. Key to success on the TOEFL. New York: Oxford University Press.
- Grellet, F. et al. 1982 Quartet, Teacher's book 1. London: Oxford University Press.

- Haile Michael Aberra. 1993. "Developing a Service English Syllabus to Meet the Academic Demands and Constraints in the Ethiopian University Context," Ph.D. Thesis. Addis Ababa: Addis Ababa University, Department of Foreign Languages and Literature. (Unpublished)
- Hanks, P. ed. 1989. The Collins School Dictionary. London: Collins.
- Harrison, A. 1983. A Language Testing Handbook. London: Macmillan Publishers.
- Heaton, B. 1975. Writing English Language Tests. London: Longman.
- _____ 1988. Writing English Language Tests, New edition. London: Longman.
- Henning, G. 1987. A Guide to Language Testing. Cambridge: Newbury House Publishers.
- Henning, G. et al. 1981. "Comprehensive Assessment of Language Proficiency and Achievement Among Learners of EFL," TESOL Quarterly, 15/4, 457-66.
- Hughes, A. 1988(a). "Achievement and Proficiency: The Missing Link," in Hughes, A. (ed.) pp.36-42.
- _____ 1988(b). "Introducing a Needs Based Test of English Language Proficiency into an English Medium University in Turkey," in Hughes, A. (ed.) pp. 134-153.
- _____ ed. 1988(c). Testing English for University Study: ELT Document 127. Hong Kong: Modern English Publications and The British Council.

- _____ 1989. Testing for Language Teachers. Cambridge: Cambridge University Press.
- Hunt, O. et al. 1990. Freshman English for Colleges, Book 1. Addis Ababa: Higher Education Main Department, Ministry of Education.
- 1990. Freshman English for Colleges, Book 2. Addis Ababa: Higher Education Main Department, Ministry of Education.
- Ingram, E. 1977. "Basic Concepts in Testing," Testing and Experimental Methods, Vol.IV, Allen, J.P.B. and Davies, A. (eds.) Oxford: Oxford University Press.
- James, K. et al. 1979 Listening Comprehension and Note-taking Course: Collins Study Skills in English. London: Collins.
- Johnson and Kin-Ling. 1981. in Pearson and Dejene.
- Kelly, R. 1978. "On the Construct Validation of Comprehension Tests: An Exercise in Applied Linguistics," Ph.D. Thesis. University of Queensland. (Unpublished)
- Kline, P. 1986. A Handbook of Test Construction. New York: Methuen.
1987. "Listing Comprehension Test 16," First Certificate in English, Cambridge: University of Cambridge Local Examinations Syndicate.
- Madsen, H.S. 1983. Techniques in Testing. Oxford: Oxford University Press.
- McEldowney, P.L. 1976. "Tests in English (overseas). The position after ten years, "Occasional Paper 36. Joint Matriculation Board, Manchester.

- McNamara, T.F. 1990. "Item Response Theory and the validation of an ESP test for health professionals," Language Testing, 7/1, 52-76.
- _____ 1991. "Test dimensionality: IRT analysis of an ESP listening test," Language Testing, 8/2, 139-159.
- Messick, S.A. 1988. " Validity," Educational Measurement, third edition. Linn, R.L. (ed.) New York: American Council on Education/ Macmillan, pp. 13-103.
- Milanovic, M. 1988. "The Construction and Validation of a Performance-based Battery of English Language Progress Tests," Ph.D. Thesis. University of London. (Unpublished)
- Ministry of Education (Ethiopia). 1982. English Curriculum (Revised),(Grades 3-8), Addis Ababa: Educational Materials Production and Distribution Agency.
- Morris, B.L. 1982. "The English Language Communicative Needs of Addis Ababa University Students," M.A. Thesis. Addis Ababa: Addis Ababa University School of Graduate Studies. (Unpublished)
- Morrow, K. 1981. "Communicative Language Testing: Revolution or Evolution?" in Alderson, J.C. and Hughes, A. (eds.) pp. 9-25.
- Munby, J. 1978. Communicative Syllabus Design. Cambridge: Cambridge University Press.
- Nuttall, C. 1982. Teaching Reading Skills in a Foreign Language. Oxford: Heinemann.
- Oller, J.W. 1979. Language Tests at School. London: Longman.

- Oller, J.W. and Jonz, J. eds. 1994. Cloze and Coherence. London: Associated University Press.
- Oller, J.W. and Perkins, K. 1980. Research in Language Testing. Rowley, Mass.: Newbury House.
- Ostler, S.E. 1980. "A Survey of Academic Needs for Advanced ESL," TESOL Quarterly, 14/4, 489-501.
- Pearson, I. 1988. "Tests as Levers for Change (or 'Putting First Things First')," Colombo: Ministry of Education.
- Pollitt, A. and Hutchinson, C. 1987. "Calibrated graded assessments: Rasch partial credit analysis of performance in writing," Language Testing, 4/1, 72-92.
1988. "Proposal for the introduction of new programme," Addis Ababa: Department of Foreign Languages and Literature, Institute of Language Studies, Addis Ababa University. (Mimeographed)
- Rees, A.L. 1980. "Reading Aloud - Suggestions for Classroom Procedure," ELTJ, 34/2, 112-121.
- Robinson, C. 1981. Themes for Proficiency. New York: Oxford University Press.
- Shaw, P. 1988. "Uses of reading aloud," Modern English Teacher, 15/4, 42-43.
- Skehan, P. 1988. "Language Testing part 1," Language Teaching. 21/4, 211-221.
- _____ 1989. "Language Testing part 2," Language Teaching. 22/1, 1-13.
- Soars, J. and Soars, L. 1987. HEADWAY. Oxford: Oxford University Press.

- Spolsky, B. 1975. "Language Testing: art or science?" Paper presented at the Fourth AILA International Congress, stuttgart.
- _____ 1973. "What does it mean to know a language, or how do you get someone to perform his competence?" Focus on the Learner. Oller, J.W. and Richards, J.C. (eds.) Rowley, Mass: Newbury house, pp. 164-176.
- Swan, M. and Walter, C. 1985. The Cambridge English Course, Student book 2. New York: Cambridge University Press.
- _____ 1987. The Cambridge English Course, Teacher's book 3. New York: Cambridge University Press.
- Underhill, N. 1982. "The Great Reliability Validity Trade off: problems in assessing the productive skills," Language Testing, Fifth in the series of Special Issues of Modern English Teacher, pp. 17-23.
- _____ 1987. Testing Spoken Language: A handbook of oral testing technique. Cambridge: Cambridge University Press.
- Vollmer, H.J. 1983. "The structure of foreign language competence," Current Developments in Language Testing. Hughes, A. and Porter, D. (eds.) Academic Press, pp.3-29.
- Walker, C. 1990. "Large-scale Oral Testing," Applied Linguistics, 11/2, 200-219.
- Wall, D. and Alderson, C. 1993. "Examining Washback: The Sri Lankan Impact Study," Language Testing, 10/1.
- Wall, D. et al. 1994. "Evaluating a placement test," Language Testing, 11/3, 321-344.

- Wallace, M.J. 1980. Study Skills in English. New York: Cambridge University Press.
- Wesche, M.B. 1987. "Second language performance testing: the Ontario Test of ESL as an example," Language Testing, 4/2, 28-47.
- Weir, C. 1988. Communicative Language Testing. Exeter: University of Exeter.
- _____ 1983. "Identifying the language problems of overseas students in tertiary education in the United Kingdom," Ph.D. Thesis. University of London.
(Unpublished)
- Woods, A. and Baker, R. 1985. "Item response theory," Language Testing, 2/2, 117-140.

APPENDICES

	<u>Page</u>
Appendix 1: Question and value distributions of Freshman English 101A and 101B (1985/86 - 1990/91) ...	209
Appendix 2: Summary of frequency of language skills ...	211
Appendix 3: Freshman English Examination (Written Test) ...	214
Appendix 4: Answer key and marking guidelines (Written Test) ...	231
Appendix 5: Freshman English Examination (Listening Test) ...	236
Appendix 6: Listening text for listening test ...	240
Appendix 7: Complete text for partial dictation ...	242
Appendix 8: Answer key and marking guidelines (Listening Test) ...	243
Appendix 9: Freshman English Examination (Oral Test set) ...	244
Appendix 10: Examiners' set for the Oral Test: source and scoring sheet, all with instructions and evaluation guidelines ...	254
Appendix 11: External written test (WT2) ...	261
Appendix 12: Answer sheet for external written test ...	284
Appendix 13: Answer key and marking guidelines for external written test ...	286
Appendix 14: Answer sheet for external listening test ...	287

Appendix 15: Answer key and marking guidelines for external listening test ...	292
Appendix 16: Student Questionnaire for the Written Test (SQ1WT) ...	293
Appendix 17: Student Questionnaire for the Listening Test (SQ2L) ...	296
Appendix 18: Student Questionnaire for the Oral Test (SQ3S) ...	299
Appendix 19: Language Teachers Questionnaire for the three tests ...	302
Appendix 20: Item results for external written test (WT2) ...	310
Appendix 21: Item results for external listening test (LT2) ...	314
Appendix 22: External listening test (Test 16), cassette (LT2) ...	316

Appendix 1

Question and value distribution of Freshman English 101A and 101B (1985/86 - 1990/91)

1985/86	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	23	24.5	15	25	18	26.1	15	30
(objective)	7	7.5	7	11.7	12	17.4	12	24
	16	17	8	13.3	6	8.7	3	6
Vocabulary (subjective)	35	37.2	20	33.3	20	29	10	20
(objective)	5	5.3	5	8.3	0	0	0	0
	30	31.9	15	25	20	29	10	20
Structure (subjective)	35	37.2	20	33.3	30	43.5	20	40
(objective)	5	5.3	5	8.3	10	14.5	10	20
	30	31.9	15	25	20	29	10	20
Composition	1	1.1	5	8.3	1	1.4	5	10
Total	94	100	60	100	69	100	50	100

1986/87	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	20	21.5	15	25	18	26.1	15	30
(objective)	10	10.8	10	16.7	12	17.4	12	24
	10	10.8	5	8.3	6	8.5	3	6
Vocabulary (subjective)	35	37.6	20	33.3	20	29	10	20
(objective)	5	5.4	5	8.3	0	0	0	0
	30	32.2	15	25	20	29	10	20
Structure (subjective)	37	39.8	20	33.3	30	43.5	20	40
(objective)	3	3.2	3	5	10	14.5	10	20
	34	36.5	17	28.3	20	29	10	20
Composition	1	1.1	5	8.3	1	1.4	5	10
Total	93	100	60	100	69	100	50	100

1987/88	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	20	21.5	15	25	20	28.2	15	30
(objective)	10	10.8	10	16.7	10	14.1	10	20
	10	10.8	5	8.3	10	14.1	5	10
Vocabulary (subjective)	35	37.6	20	33.3	20	28.2	10	20
(objective)	5	5.4	5	8.3	0	0	0	0
	30	32.2	15	25	20	28.1	10	20
Structure (subjective)	37	39.8	20	33.3	30	42.2	20	40
(objective)	3	3.2	3	5	10	14.1	10	20
	34	36.5	17	28.3	20	28.2	10	20
Composition	1	1.1	5	8.3	1	1.4	5	10
Total	93	100	60	100	71	100	50	100

1988/89	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	20	21.5	15	25	20	32.3	15	33.3
(objective)	10	10.8	10	16.7	10	16.1	10	22.2
	10	10.8	5	8.3	10	16.1	5	11.1
Vocabulary (subjective)	35	37.6	20	33.3	20	32.3	10	22.2
(objective)	5	5.4	5	8.3	0	0	0	0
	30	32.2	15	25	20	32.3	10	22.2
Structure (subjective)	37	39.8	20	33.3	21	33.8	15	33.3
(objective)	3	3.2	3	5	9	14.5	9	20
	34	36.5	17	28.3	12	19.4	6	13.3
Composition	1	1.1	5	8.3	1	1.6	5	11.1
Total	93	100	60	100	62	100	45	100

1989/90	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	23	22.2	15	27.3	24	33.3	17	34
(objective)	7	8.6	7	12.7	10	13.9	10	20
	16	13.6	8	14.5	14	19.4	7	14
Vocabulary (subjective)	20	24.7	10	18.2	20	27.8	10	20
(objective)	0	0	0	0	0	0	0	0
	20	24.7	10	18.2	20	27.8	10	20
Structure (subjective)	37	45.7	20	36.3	26	36.1	13	26
(objective)	3	3.7	3	5.5	0	0	0	0
	34	42	17	30.9	26	36.1	13	26
Composition	2:1+5	7.4	10	18.2	2	2.8	10	20
Total	86	100	55	100	72	100	50	100

1990/91	101A				101B			
	Qn/y	%	Val	%	Qn/y	%	Val	%
Comprehension (subjective)	22	29.3	15	30	31	37.4	20	36.4
(objective)	8	10.7	8	16	9	10.8	9	16.4
	14	18.7	7	14	22	26.5	11	20
Vocabulary (subjective)	20	26.7	10	20	24	28.9	12	21.8
(objective)	0	0	0	0	0	0	0	0
	20	26.7	10	20	24	28.9	12	21.8
Structure (subjective)	27	36	15	30	26	31.3	13	23.6
(objective)	3	4	3	6	0	0	0	0
	24	32	12	24	26	31.3	13	23.6
Composition	2:1+5	8	10	20	2	2.4	10	18.2
Total	75	100	50	100	83	100	55	100

Abbreviations: Qn/y = Questions, Quantity

Val = Value or marks

Appendix 2

Summary of frequency of language skills

<u>Language skill</u>	<u>Reading</u>	<u>Language Activity in:</u>			<u>Total</u>
		<u>Listening</u>	<u>Writing</u>	<u>Speaking</u>	
39	2,3	1,4	1,2	1	7
51	2,3,4	-	1,2	1	6
26	1	3,4	1,2	-	5
32	1	1,2	2	-	4
53	-	1	1	1	3
52	2,3	-	1,2	-	4
29	-	-	1,2	2	3
35	2,3	1,4	-	-	4
37	2,3	1,4	-	-	4
38	-	-	2	1,2	3
20	1,4	2	-	-	3
28	1	2,3	-	-	3
30	1	1,2	-	-	3
36	-	-	2	1	2
40	2,3	1	-	-	3
44	2,4	-	1	-	3
50	-	-	2	1	2
24	1	2	-	-	2
43	-	1	1	-	2
46	1,2,3,4	-	-	-	4
17	1,4	-	-	-	2
19	1,4	-	-	-	2
41	2,3	-	-	-	2
45	2,3	-	-	-	2
11	-	1,2,3,4	-	-	4

13	-	3	-	-	1
18	-	-	1,2	-	2
21	-	-	1,2	-	2
25	-	-	1,2	-	2
23	-	-	2	-	1
31	-	-	2	-	1
12	-	-	-	1,2	2
27	-	-	-	1,2	2
47	-	-	-	1,2	2
48	-	-	-	1,2	2
49	-	-	-	1,2	2
4	-	-	-	1	1
14	-	-	-	2	1
54	-	-	-	2	1
<u>Total</u>	<u>34</u>	<u>24</u>	<u>25</u>	<u>21</u>	<u>104</u>

NOTE: The numbers ranging from 1-4 in the macro-skill columns refer to language activities, which are given below for each macro-skill.

Reading

1 = reading instructions

2 = reading textbooks

3 = reading handouts

4 = using the dictionary

Listening

1 = listening to lectures

2 = listening to instructions/
explanations

3 = listening to teachers'
questions

4 = listening to discussions

Writing

1 = taking notes from
lectures or books

2 = short or essay
examination answers

Speaking

1 = answering questions

2 = asking questions

Appendix 3

Freshman English Examination (Written Test)

FRESHMAN ENGLISH EXAMINATION

(WRITTEN TEST)

STUDENT'S NAME: _____

I.D.NO: _____

FACULTY: _____

DEPARTMENT: _____

PROGRAMME (DEGREE/DIPLOMA): _____

SECTION: _____

GENERAL DIRECTION:

This test booklet contains 13 sections (i.e, A -M). You have 3 hours to attempt all the questions and tasks according to the instructions for each section.

Reading Passage for Sections A - D

Read the following passage and use the information to respond to the questions in sections A - D

1. Africa has exceptionally great water-power resources together with some resources for thermal power production. Power production has increased at an impressive rate over the past decades, and is
 4 continuing to rise very rapidly. Some degree of limitation is imposed by shortages of local capital and skilled manpower, but the main factor governing the pace of development seems to be the level of demand for electricity. This clearly indicates the close relationship which exists between power production and other forms of economic
 9 activity. While the provision of electricity is not a sufficient condition for rapid economic development, it can greatly assist the process, and it is certainly a necessary condition for industrial growth.

2. The power resources of Africa are not only ample but also
 14 wide-spread, and so it has been possible for new developments in this field to be widely distributed. Nevertheless, development on a really large scale has been concentrated at only a few points where a number of factors have been particularly favourable. Kariba on the Zambezi in Zimbabwe, Akosombo on the Volta in Ghana and
 19 Kanji on the Niger in Nigeria all offered suitable sites for damming a major river to create a great head of water, and at the same time that each decision to proceed was taken these countries were each relatively strong economically and politically. Each government was able to contribute part of the capital cost and to
 24 obtain the remainder of the necessary finance from overseas. The Inga project in Zaire offers an even more suitable physical site, but Zaire did not until recently have the political stability or the economic strength to attract the capital required to proceed with the project.

28 3. Zambia and Zimbabwe together provided a large market for power compared with most parts of Africa, and Nigeria provides sufficient demand to justify a major scheme. The relative prosperity of Ghana creates a market larger than in most countries of its size, but a

- 2 -

large aluminium project was also needed to support the volta scheme.
 33 Many countries in Africa do not have the market to justify such
 schemes.

4. There are some countries, and especially some cities and mining
 areas, with only a limited hydro-electric potential near at hand,
 where level of demand has justified investment in thermal power
 38 stations, generally using imported fuel. Now every effort is being
 made to avoid this policy wherever possible, except where local oil
 or gas resources exist. In some cases the need for it may be
 avoided by greater international co-operation within Africa. This
 could also prevent duplication of expenditure on hydroelectric
 43 plants in neighbouring countries, encouraging concentration upon
 those sites where power can be generated at lowest cost. Until
 recently, except at Kariba, transmission across frontiers has
 generally taken place only where internal demand has risen more
 slowly than was hoped, as in the case of sales from Uganda to Kenya,
 48 and from Ghana to Togo and Benin. However, it could be of great
 value elsewhere, as at Inga, reducing costs for both the exporting
 and the importing countries. At the same time, progress is being made
 on developing the transmission networks within countries to form
 52 national grids, which both enable the lowest cost power stations to
 serve the largest possible areas and reduce the disruptions created
 when breakdowns occur.

Section A

Decide in which paragraph (ie., 1,2,3,or 4) each of the
 following statements (items 1 & 2) are contained, and write
 your answer in the space provided.

1. Power resources are abundantly widespread in Africa.

1. Par. No. _____

2. Transmission across frontiers could benefit both the exporting
 and the importing countries by reducing costs.

2. Par.No. _____

...#

- 3 -

3. Choose the most suitable title for the passage, and write the letter of your choice in the space provided.
- a) Progress in the supply of power in Africa.
 - b) Problems in water-power resources in Africa.
 - c) The supply of power in Africa: Progress and problems.
 - d) Problems in the supply of power in Africa.

3. _____

Section B

Provide brief answers (notes) for the following statements according to information in the passage.

4. At least three factors are necessary to develop the power resources of Africa; a) _____
 b) _____
 c) _____
5. At least three favourable factors enabled Kariba, Akosombo and Kanji to be developed:
 a) _____
 b) _____
 c) _____
6. African countries with large enough markets to justify large power projects: a) _____
 b) _____
 c) _____
7. The main disadvantage of thermal power stations:
 a) _____

...#

- 4 -

Section C

Some words and phrases are used by writers to link ideas within the same sentence and between sentences to achieve a smooth flow down the page. Decide whether each statement is True or False depending on how the words or phrases are used in the passage.

8. "While ..." in par. 1, line 9 shows time relationship of similar ideas between sentences. 8. _____
9. "... and ..." in par. 3, line 29 is used to give an example within the same sentence. 9. _____
10. "This ..." in par. 4, line 41 refers to international co-operation within Africa. 10. _____
11. "Until ..." in par. 4, line 44 shows time relationship of ideas between sentences. 11. _____
12. "... as ..." in par. 4, line 49 is used to give an example within the same sentence. 12. _____

Section D

The words under column A are used in the passage. Match the words in column A with the meanings under column B, and write the letter of your answer in the space provided.

<u>Column A</u>	<u>Column B</u>
_____ 13. impressive	a) a period of ten years
_____ 14. pace	b) enough or more than enough
_____ 15. ample	c) the rate or speed of something
_____ 16. potential	d) causing admiration by giving; a feeling of size and/or importance
_____ 17. near at hand	e) to force something on someone
	f) close
	g) existing in possibility; not at present active or developed, but able to become so
	h) disorder

.../

Section E

Library catalogue cards

You have cards from a library catalogue with information on books. Using information from this source write the answers for the following questions in the space provided.

418.00 Rivers, W.M..Speaking in many tongues/
RIV Wilga M.Rivers. ___3rd ed. ___ Cambridge:
Cambridge University Press, 1983. ___ 258p
24 cm
Includes index

428.2 Burton, S.H..Mastering English Grammar/
BHR S.H. Burton. ___ London:Macmillan,
1984. ___178p; 23 cm

428.24 Rinvoluceri, M..Grammar games/
RIN Mario Rinvoluceri. ___ Cambridge:
Cambridge University Press, 1984. ___
138; 25 cm

808.04 Johnson, K..Communicating in writing/
JOH Keith Johnson. ___ Harlow: Longman,
1981. ___125p, 24 cm

18. Write the second (sur)names of the authors (writers) in alphabetical order.

a) _____

c) _____

b) _____

d) _____

19. What is the title of the book written by Rinvoluceri?

19. _____

20. What is the call number for the book written by Rivers?

20. _____

21. Where (city) is the book written by Burton published?

21. _____

22. When was Johnson's book published?

22. _____

.../

fain

action, or event, e.g. *The meeting was a complete failure.*

3 Your failure to do something is not doing something that you were expected to do, e.g. *His friends remarked on his failure to appear at the party.*

Similar words: (sense 2) flop, washout

fain

(adverb; an old-fashioned word) willingly or gladly.

[from Old English *fægen* meaning 'glad' or 'happy']

faint, fainter, faintest; faints, fainting, fainted

1 (adjective) A sound, colour, or feeling that is faint has little strength or intensity.

2 (verb) If you faint, you lose consciousness for a short time.

3 (adjective) If you feel faint, you feel weak, dizzy, and unsteady.

faintly (adverb), faintness (noun).

[from Old French *feint* meaning 'weak' or 'sluggish']

Similar words: (sense 2) pass out, black out, swoon, keel over

faint-hearted

(adjective) lacking courage and confidence.

fair, fairer, fairest; fairs

1 (adjective) reasonable or equal according to generally accepted ideas about what is right and just, e.g. *That's not fair!*

2 quite large, e.g. *She made a fair amount of money.*

3 moderately good or likely to be correct, e.g. *I had a pretty fair idea of the answer.*

4 having light coloured hair or pale skin.

5 (an old-fashioned use) attractive to look at, e.g. *this fair city of ours.*

6 with pleasant, dry, and fine weather, e.g. *It promises to be fair.*

7 (noun) A fair is a form of open-air entertainment with stalls, sideshows, and machines to ride on.

8 A fair is also an exhibition of goods produced by a particular industry, e.g. *the Frankfurt Book Fair.*

fairly (adverb), fairness (noun)

[Senses 1-6 are from Old English *fæger* meaning 'lovely'; senses 7-8 are from Latin *feria* meaning 'festival']

Similar words: (sense 1) just, impartial, equitable, unbiased

fairway, fairways

(noun) On a golf course, the fairway is the area of trimmed grass between a tee and a green.

fairy, fairies

(noun) In stories, fairies are small, supernatural creatures with magical powers.

[from Old French *féerie* meaning 'fairyland']

fairy lights

(plural noun) Fairy lights are small, coloured

electric lights used for decoration, for example on Christmas trees.

fairy tale, fairy tales

(noun) A fairy tale is a story of magical events.

fait accompli, faits accomplis (pronounced fate ak-kom-plee)

(noun) A fait accompli is something that has already been done and cannot be changed.

[a French expression meaning 'completed deed']

faith, faiths

1 (noun) Faith is a feeling of confidence, trust or optimism about something.

2 A faith is a particular religion.

3 (phrase) If you do something in good faith, your reasons are honest and sincere.

[from Latin *fides* meaning 'trust']

Similar words: (sense 1) confidence, conviction, trust

faithful

1 (adjective) loyal to someone or something and remaining firm in support of them.

2 accurate and truthful, e.g. *The film was faithful to the original book.*

faithfully (adverb), faithfulness (noun).

Similar words: (sense 1) steadfast, loyal, staunch

faith healing

(noun) Faith healing is the treatment of a sick person through prayer and the power of religious faith.

faith healer (noun).

faithless

(adjective) disloyal or dishonest.

fake, fakes, faking, faked

1 (noun) A fake is an imitation of something made to trick people into thinking that it is genuine.

2 (adjective) Fake means imitation and not genuine, e.g. *fake antiques.*

3 (verb) If you fake an emotion or feeling, you pretend that you are experiencing it.

Similar words: (sense 1) forgery, phoney, fraud, sham (sense 2) phony, bogus, counterfeit, fraudulent, sham, false (sense 3) counterfeit

falcon, falcons

(noun) A falcon is a bird of prey that can be trained to hunt other birds or small animals.

[from Latin *falco* meaning 'hawk']

fall, falls, falling, fell, fallen

1 (verb) If someone or something falls or falls over, they drop towards the ground.

2 (noun) If you have a fall, you accidentally fall over.

3 A fall of snow, soot, or other substance is a quantity of it that has fallen to the ground.

4 (verb) If something falls somewhere, it lands there, e.g. *A shadow fell over her book.*

5 If something falls in amount or strength, it becomes less or weaker, e.g. *Oil consumption fell by 24%.*

- 7 -

Section F

On page 6 you find a page from a dictionary with information on words. Using information from this source write the answers for the following questions in the space provided.

23. Give a one word meaning of the word "fain".
23. _____
24. Give the noun form of the word "faint".
24. _____
25. The word "falcon" comes (was introduced in English) from which other language?
25. _____
26. Give the plural form of the word "fairy".
26. _____
27. Give an example of the word "fake" when used as an adjective.
27. _____

Section G

Your role is defined below. Page 8 is part of a course offering with information on the courses you take. Based on this source of information and the role you are given, complete the registration form on page 9.

Your Role

You are a second year student in the Department of Foreign Languages and Literature (FL & Lit.) in the Institute of Language Studies (ILS). Your minor subject is Geography. You have completed the first semester of your second year in the University, and you are now registering for the second semester of the same year as an English major and Geography minor.

Now complete the registration form on page 9

.../

- 8 -

ADDIS ABABA UNIVERSITY
INSTITUTE OF LANGUAGES STUDIES
DEPARTMENT OF FOREIGN LANGUAGES AND LITERATURE
SECOND SEMESTER 1992/93
COURSE OFFERING-UNDERGRADUATE PROGRAMME

Year II: ETLA & Lit. Minors

<u>Course No.</u>	<u>Course Title</u>	<u>Credit</u>	<u>Instructor</u>
FLEn 202	Intermediate English I	3	Mammo K.
FLEn 204	Spoken English II	3	Getachew F.
Ling 202	Int. to Lang. & Ling. II	3	Aklilu
Educ 211	Principles of Curr. Enquiry	4	T.B.A
ELAm 204	Advanced Comp. II	3	Yitbarek
Total		16	

Year II: Geography Minors

FLEn 202	Intermediate English I	3	Mammo/Sharma
FLEn 204	Spoken English II	3	Getachew F.
Ling 202	Int. to Lang. & Ling. II	3	Aklilu
Educ 211	Principles of Curr. Enquiry	3	T.B.A
Geog 222	Int. to Geography	3	T.B.A
Geog 212	Int. to Econ. Geography	3	T.B.A
Total		18	

Year III: Geography Minors

FLEn 302	Advanced Comp. I	3	Italo
FLEn 306	Skills Dev. Methodology II	4	Atkins
EDUC 312	General Methods of Teaching	3	T.B.A
EdAd 302	Survey of Org. & Management in Ethiopian Schools	2	T.B.A
Ling 314	English Morph. & Syntax	3	Baye Y.
Geog 232	Geography of Africa	3	T.B.A
Total		18	

.../

Section H

In the passage below, certain words have been deleted or removed. Choose the best word from the list given below to fill in the blanks, and write your answer in the space provided. Use one word only once.

decrease	surviving	growing	did	take
are	a	survive	increases	the
does	was	will	grow	is

The wealth or level of development of a country is usually measured by the Gross National Product (GNP) of that country. GNP is the market value of all goods and services produced by the people of a country over a given time (usually a year). GNP is often divided by the number of people in a country to give an average figure called Gross National Product per head or per capita.

Because GNP is an average figure, it 28 not show the 28. _____
 difference in wealth between the richest and poorest people
 in a country. For example, in Brazil in 1973 the GNP per
 capita was \$750 but 18 million people were 29 on less 29. _____
 than \$75 per year. In most countries there are also
 differences in wealth between one region and another. But
 a country's GNP does not 30 these regional differences 30. _____
 into account. 31 large share of the national wealth may 31. _____
 be concentrated in the region containing the capital city.
 GNP for the whole country may 32 but the differences 32. _____
 between the capital city and the rest of the country can
 have actually widened. This is characteristic of many poor
 countries. On the one hand, there are some regions which
 are quite wealthy, on the other, there are regions which are
 very poor.

Growth in GNP per capita 33 a major aim of many 33. _____
 countries. However, if the population of a country 34 at 34. _____
 a higher rate than the GNP, some people 35 not be better 35. _____
 off.

The world's population is 36 very fast and it is 36. _____
 growing most rapidly in the poorer countries. As we have
 seen, if the rate of population growth in a country is higher
 than the rate of economic growth, the GNP per capita will
37 . This means that poor countries have to make great 37. _____
 efforts just to maintain GNP per capita at the same level.

.../

Section I

In the short passage below you are given a choice of words/ phrases in brackets to fill the gap. Choose the most suitable word(s) to fill in the gap, and write your answer in the space provided.

Such is the demand for oil in the world that a country suddenly becomes richer if it has large quantities of oil. The large oil companies carry out the work, (38. so that, but, this) they must obtain licences from the Government in order to do so and they must pay for these licences. 30. _____

More important than these licences is the money or revenue collected from the sale of oil. The Government takes at least half the money from the sale of the oil. This provides much needed money for the development of the country. A few years ago Nigeria made primary education free for all children and soon afterwards made it compulsory. Without money from oil, this would not have been possible. Oil revenues have also helped in building hospitals, roads and irrigation works and in agricultural development.

In addition, the oil industry provides work for many people. They spend their wages on consumer goods, (39. thus, this, as) encouraging other trades and services to develop. 39. _____

It is important that a country uses its oil revenues wisely (40. so that, despite, this) the results can be felt in the future, as the oil will not last forever. It is doubtful if Nigeria will be producing at this level in twenty five years time. 40. _____

(41. In addition, Despite, But) being the leading producer in Africa, Nigeria's oil revenues have to be used for many people, (42. this, but, as) it is also the most populous country in Africa. We refer to such production compared with population as per capita production, that is, per head of the population. 41. _____
42. _____

.../

Section J

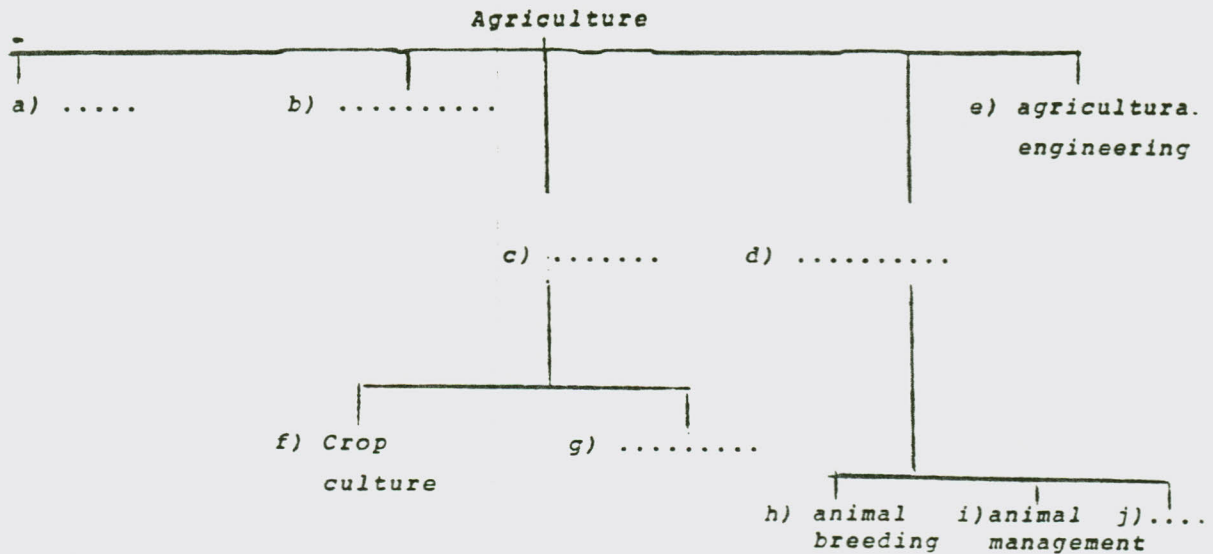
Read the following paragraph and answer the questions on it.

Although farming is no doubt the most important part of agriculture, agriculture includes other activities in addition to farming. Agriculture has several major components such as agricultural economics, which deals with the business aspects of agriculture. Agronomy is the branch of agriculture which is concerned with crop culture and soil management, and animal science includes animal breeding, animal management and biochemistry. Another important branch of agriculture is agricultural engineering.

43. Decide which of the following statements contains the main idea of the paragraph, and write the letter of your choice in the space provided.

- a) It describes the various branches of farming.
 - b) It defines the various branches of agriculture
 - c) It describes the various branches of agriculture
 - d) It classifies the various activities of farming
43. _____

44. Use information in the paragraph to complete the following tree- diagram.



Section L

The following extract contains a number of errors in grammar, spelling and punctuation. First underline the errors, then correct the errors in the space provided under each line by writing only the necessary corrections. Do not re-write the passage in your own words.

Example: I am an student of English *
English

46. Agricultural Education

education has always been consider as an important factor in
achieving economic growth and development and human progress.

It plays a significant role in acclerating agricultural
development in the early stages of economic development.

Education is the acquisition and application of knowledge and the
use of new ideas can lead to a change in the future behaviour
patterns of man Education and training provide man with useful
skills who can be applied to improve his economic performance.

They can also used to improve the quality of labour and
management and reduce tradionalism, thereby encouraging the
adoption of new ideas.

Section M

In the passage below some single words have been deleted or removed.
As you read the passage find appropriate single words to fill in the
blanks, and write your answers in the space provided.

By the end of the 1970s more than half the
world's city dwellers were living in the developing
world. By the year 2000 more than 60 % of the world's
city dwellers will be living in developing countries.

.../

- 15 -

In the industrialised countries, populations 47 47. _____
 generally rising only slowly and 48 some cases are 48. _____
 static. In the cities 49 developing countries 49. _____
 populations are increasing to a 50 high level. This is 50. _____
 the result 51 high birth rates and low death 52 . 51. _____
 52. _____
 The poorest urban dwellers do suffer 53 periodic 53. _____
 outbreaks of infectious diseases, but 54 general 54. _____
 fewer babies die today and people 55 longer. 55. _____

In addition to natural 56 , movement of 56. _____
 people from the countryside 57 towns adds to the 57. _____
 growth of 58 in developing countries. In some 58. _____
 cities 59 has accounted for between 50% 60 59. _____
 60. _____
 70% of their population growth in 61 last 20 years. 61. _____

One very important 62 between the urban 62. _____
 growth in the 63 world in the 19th century and 63. _____
64 growth in developing countries at the 65 64. _____
 65. _____
 time is that Europe in the 66 century it was 66. _____
 associated with rapid 67 . There was a shortage of 67. _____
 labour 68 factories, mines and shipyards. This is 68. _____
69 generally the case in developing countries 69. _____
70 . Urban growth is largely caused by 71 70. _____
 71. _____
 of development in the countryside rather 72 72. _____
 increased opportunities in the cities. As 73 73. _____
 result, there are more people in 74 cities than 74. _____
 there are jobs. Yet 75 massive movement from the 75. _____
 countryside, mainly 76 young people, continues 76. _____
 and looks likely 77 continue in the future. 77. _____

What finally 78 a person to leave his 78. _____
 village? 79 may be dissatisfaction with village 79. _____
 life, 80 the feeling that there are greater 81 80. _____
 81. _____
 in the city. Most of the 82 deciding to make a 82. _____
 fresh start 83 the city are likely to 84 83. _____
 84. _____
 undernourished, poor and unemployed.

.../

- 16 -

Under-nourishment or malnutrition 85
be the result of not having 86 land on which
to grow crops, or it may be because of the lack
of opportunities to work in order to supplement
the farm income. Also, perhaps surprisingly, it
may be the result of the modernisation of farming
methods.

85. _____

86. _____

WELL DONE !

2

- 20. 410.00 RIV
- 21. London
- 22. 1991

F) 1 mark each (5 marks)

- 23. Willingly/gladly
- 24. faintness
- 25. Latin
- 26. fairies
- 27. fake antiques

.../

3

- G) A quarter (1/4) mark for each blank (5 marks); no penalty for spelling, capitalization and punctuation errors. There are 20 blanks excluding ones for Age, Nationality, Sex, Advisor, Registrar and the specific course credits.

ADDIS ABABA UNIVERSITY
OFFICE OF THE REGISTRAR

Age ignore (1) YEAR II/2/two (2) SEMESTER II/2 two
(3) 19 92/93 1985/ ignore

(4) See front page of question paper
NAME IN BLOCK LETTERS

(5) IS RE_ENROLLED IN THE FACULTY OF I.L.S. (in full)

(6) see front page of paper ignore ignore (7) FL&Lit(in fu.
I.D.No, NATIONALITY SEX DEPARTMENT

<u>COURSE TITLE</u>	<u>COURSE NO.</u>	<u>CREDIT</u>
(8) <u>Intermediate English I</u>	(14) <u>FLEn 202</u>	<u>3 ignore</u>
(9) <u>Spoken English II</u>	(15) <u>FLEn 204</u>	<u>3 "</u>
(10) <u>Int. to Lang. & Ling. II</u>	(16) <u>LING 202</u>	<u>3 "</u>
(11) <u>Principles of Curr. Enquire</u>	(17) <u>Educ 211</u>	<u>3 "</u>
(12) <u>Int. to Geography</u>	(18) <u>Geoc 222</u>	<u>3 "</u>
(13) <u>Int. to Econ. Geography</u>	(19) <u>Geoc. 212</u>	<u>3 "</u>
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
	(20) TOTAL	<u>18</u>

ignore
ADVISOR

ignore
REGISTRAR

.../

4

H) 1Mark each (10 marks)

- | | |
|---------------|---------------|
| 28. does | 33. is |
| 29. surviving | 34. increases |
| 30. take | 35. will |
| 31. a | 36. growing |
| 32. grow | 37. decrease |

I) 1 mark each (5 marks)

- | | |
|-------------|-------------|
| 38. but | 41. despite |
| 39. thus | 42. as |
| 40. so that | |

J) 1 mark each (7 marks) ; no penalty for spelling errors.

43. C

- | | |
|---------------------------|--------------------|
| a) farming | d) animal science |
| b) agricultural economics | e) soil management |
| c) agronomy | f) biochemistry |

K) 5 marks

44. consider organisation of content and language.

* Marking guidelines for the paragraph

5=Very comprehensible and mostly free from language errors.

4=Reasonably comprehensible with a few language errors.

3=Comprehensible but with many language errors.

2=Partially comprehensible with too many language errors.

1/0=Mostly incomprehensible

* A possible model, for example, for the paragraph.

How formal education is organised in Ethiopia

Formal education in Ethiopia, as in other countries, is divided into three sectors. These sectors are primary, secondary and tertiary. There are about 7,500 primary schools in the country. senior secondary schools and technical and vocational schools. There are about 800 secondary schools in Ethiopia. The tertiary sector consists of universities, professional colleges, teacher training institutes, teacher training colleges and a polytechnic institute. The professional colleges can be further subdivided into animal health, agricultural, commercial and forestry colleges.

.../

5

L) 1 mark each (5 marks)

46) line 1 : Education line 2: ... who → which
 line 1 : considered line 3: can also be used
 line 7: ...man.

M) 1 mark each (40 marks); accept only exact words given here, and no penalty for spelling errors.

47. are	57. to	67. industrialisation	77. to
48. in	58. cities/ population	68. for/in	78. causes
49. of	59. migration	69. not	79. it
50. very	60. and	70. today	80. or
51. of	61. the	71. lack	81. opport. ities
52. rates	62. difference	72. than	82. people
53. from	63. industrialised	73. a	83. in
54. in	64. urban	74. the	84. be
55. live	65. present	75. the	85. may
56. growth	66. 19th	76. of	86. enough

.../

Appendix 5

Freshman English Examination (Listening Test)

FRESHMAN ENGLISH EXAMINATION

(Listening Test)

Student's Name: _____
I.D.No.: _____
Faculty: _____
Department: _____
Programme (Degree/Diploma): _____
Section: _____

General Direction:

This test booklet contains 3 sections (ie, A-C). You have 30 minutes to attempt all the questions and tasks according to the instructions for each section.

- 1 -

Section A Partial Dictation

You have a text of a talk with one or more words missing. The examiner is going to read the full (complete) text twice. The first time, he will read it slowly and he will pause frequently for you to write down the missing word (s). The second time, he will read it straight through again for you to check what you have written. Now complete the talk by writing one or more words in each space as you listen to the talk.

1. When a student takes notes in a lecture _____ do four
2. things. Firstly, he has to understand _____, as the
3. speaker says it. He _____ the lecture in order to look
4. up a new word or check an unfamiliar sentence pattern.
5. The second thing the student has to do is to decide what's
6. important in the lecture. Often the lecturer _____.
7. He may do this directly or indirectly. If he says 'This
8. next point is important', the student will have little difficulty.
9. The lecturer may, however, use a more colloquial style. A
10. sentence such as "This is the crunch", meaning _____ the
11. really important point, will often cause difficulty. Many
12. lecturers pause, and speak more slowly and loudly,
13. _____ making an important point. If the student can't
14. recognize these indirect signals, _____ it difficult to
15. decide what's important.
16. The third _____ the student is that he has to write
17. down the important points. There are two problems here in addition
18. to _____ important. The first one is speed. The
19. second one is clarity. The student should abbreviate, should
20. write down the important information words (usually nouns, sometimes
21. verbs or adjectives) and should write one point on each line.
22. _____, if possible, _____ a moment to write when
23. the lecturer _____ vital information.
24. Finally, the student's notes must show the connections
25. between the various points _____. If he makes intelligent
26. use of spacing and underlining, together with the employment of
27. conventional _____ and the numbering of points _____
28. understand the framework of the lecture more easily.

.../

Direction for sections B and C

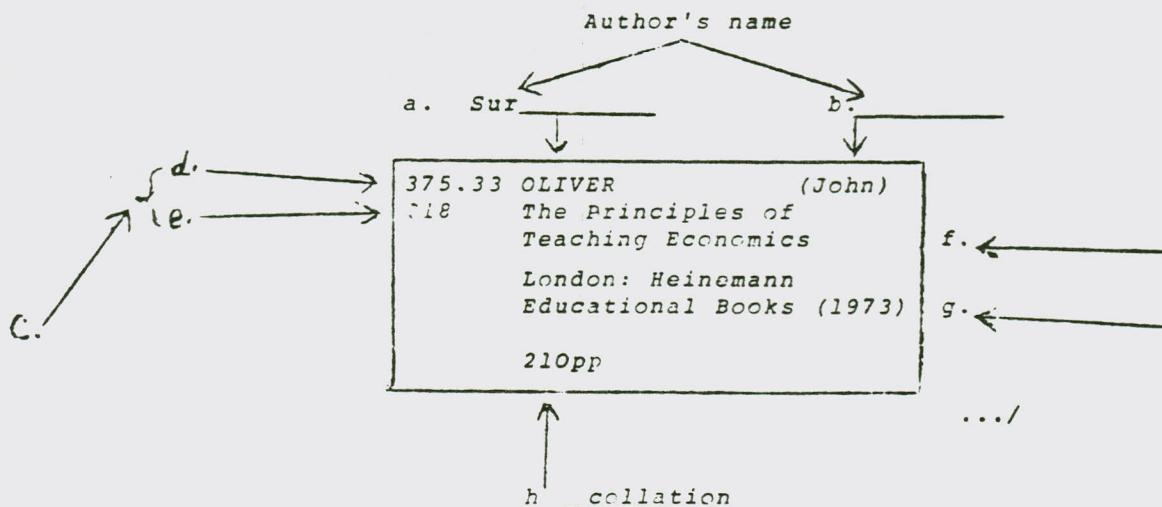
You are now going to listen to a talk. The listening passage will be read out twice. First, study the questions on your question paper. Then, at the end of the first reading you have five minutes to attempt to answer some of the questions in section B and C. Finally, at the end of the second reading you have another five minutes to attempt the rest of the questions.

Section B

1. Decide which of the following is the most appropriate title for the listening passage, and write the letter of your choice in the space provided.
 - a) using references
 - b) borrowing books
 - c) using a library
 - d) locating books

1. _____

2. The diagram below is a sample library card. Label the parts as you listen to the passage.



Section C

Choose from the set of choices given in the box to answer the questions below. Use one choice only once, and write your answer in the space provided.

-
- | | |
|--------------------|-------------------------|
| . name of author | . old editions of books |
| . borrowing slip | . the catalogues |
| . dictionaries | . title of book |
| . author catalogue | . bibliography |
| . encyclopedia | . periodicals |
-

3. What two things should a student know to borrow a book from a library?

a) _____

b) _____

4. What are examples of the three types of materials kept in the reference section of a library?

c. _____

d. _____

e. _____

5. What consists of a list of books entered on cards in alphabetical order under the name of the author, institution or editor?

f. _____

6. What type of a book gives an overview of a topic, including definition, description, background and etc.?

g. _____

WELL DONE!

Appendix 6

Listening text for listening test

LISTENING TEXTDirection for sections B and C

You are now going to listen to a talk. The listening passage will be read out twice. At the end of the first reading you have five minutes to attempt to answer some of the questions in sections B and C. Finally, at the end of the second reading you have another five minutes to attempt the rest of the questions. Ready?

The uses a student makes of his college library may be summarized under three broad headings: borrowing books, consulting reference materials and general study.

First, then, borrowing books. Let's assume a student has been given the author and title of a certain book which he's been told to read, and he wants to borrow it from the library. As soon as he gets to the library he should check that the book is in stock, make sure that it's available for lending, and find out where to locate it on the library shelves. In order to do this he'll have to consult the catalogues. As he has the name of the author, it'll be easiest to use the author, or name, catalogue. This consists of a list of books entered on cards in alphabetical order under the name of the author, institution or editor. If you look at your question paper you'll see a sample card, the most important item in the whole business of using a library.

At the top of the card, then, in the centre, is written the author's name-- his surname always comes first, followed by his forename. The forename is further distinguished from the surname by punctuation (brackets, or a comma).

The book's call number, which identifies the book, occurs near the top left hand corner of the card. This number normally consists of two parts, each part being printed on a separate line. The first part, on the top line, is the class number, which tells you in which subject area the book lies. The Dewey Decimal System, which nearly all libraries use to classify their books, divides all knowledge up into major subject areas. Every subject has a number and the number after the decimal point refers to a particular branch of that subject. Thus the sample card shows the class number as 375.33. Beneath this Dewey Decimal number you'll see another number for the author. The capital letter O represents the first letter of the author's surname. The following number is especially assigned to this particular name.

.../

- 2 -

While the class number tells the student which general area in the library to go to in order to find the book, the author number will direct him to the exact shelf.

Now before we move on to consulting reference material, we'll just briefly examine the other details on the name card from the catalogue. Under the author's name, the title of the book is printed. There then appears the imprint details: these give details of publisher, place and date of publication. Finally, under the imprint, you have the collation details. These give number of volumes, pages, illustrations and maps.

But now let's look at the second broad use: consulting reference material. In addition to the lending section an academic library always has a reference section. Three important types of materials in this section are:

- (1) books which were planned and written to be referred to for pieces of information rather than to be read completely. Examples of such books included dictionaries, bibliographies and encyclopedias.
- (2) books which are valuable or difficult to replace. An example would be an old edition of a book, of which the library has only a single copy, and which is out-of-print.
- (3) professional journals or periodicals.

A student who's starting on a new and difficult topic often finds it valuable to consult an encyclopedia. This type of book is arranged alphabetically by subjects and it gives an overview of a topic, including definition, description, background and etc. When the student is writing a dissertation or doing a piece of research then he will need to consult a specialized bibliography. Very recent research, however, may not appear in a bibliography and the students should, therefore, always consult the relevant professional journals.

On many occasions, however, the student may well come to the library because it provides a suitable working environment. Moreover, he can, if he meets an unexpected problem in his work consult one of the reference books or perhaps even a fellow-student. He can also have a chat with his fellow-students in his breaks.

..#

Appendix 7

Complete text for partial dictation

Section A partial Dictation

You have a text of a talk with one or more words missing. I am going to read the full (complete) text twice. The first time, I will read it slowly and I will pause frequently for you to write down the missing word(s). The second time I will read it straight through again for you to check what you have written. Now complete the talk by writing one or more words in each space as you listen to the talk. Ready?

1. When a student takes notes in a lecture he has to do four things.
2. Firstly, he has to understand what's said, as the speaker says it.
3. He can't stop the lecture in order to look up a new word or check
4. an unfamiliar sentence pattern.
5. The second thing the student has to do is to decide what's
6. important in the lecture. Often the lecturer signals this. He
7. may do this directly or indirectly. If he says 'This next point
8. is important', the student will have little difficulty. The
9. lecturer may, however, use a more colloquial style. A sentence
10. such as 'This is the crunch', meaning this is the really
11. important point, will often cause difficulty. Many lecturers
12. pause, and speak more slowly and loudly, when they are making
13. an important point. If the student can't recognize these indirect
14. signals, he will find it difficult to decide what's important.
15. The third task that faces the student is that he has to
16. write down the important points. There are two problems here
17. in addition to deciding what's important. The first one is
18. speed. The second one is clarity. The student should abbreviate,
19. should write down the important information words (usually nouns,
20. sometimes verbs or adjectives) and should write one point on
21. each line. He ought, if possible, to choose a moment to write
22. when the lecturer is not giving vital information.
23. Finally, the student's notes must show the connections
24. between the various points he has noted. If he makes intelligent
25. use of spacing and underlining, together with the employment
26. of conventional signs and the numbering of points, he will be
27. able to understand the framework of the lecture more easily.

.../

Appendix 8

Answer key and marking guidelines (Listening Test)

Freshman English Examination

*Answer Key and marking guidelines

II. Listening test

- A) 1 mark for each blank (15 marks); no penalty for capitalization and spelling errors as long as meaning is not affected.

Line No 1. he has to	Line No 12/13, when they are	Line No 22, to choos.
" " 2. what's said	" " 14, he will find	" " 23, is not giving
" " 3. Can't stop	" " 16. task that faces	" " 25, he has noted
" " 6. signals this	" " 18. deciding what's	" " 27, signs,
" "10. this is	" " 22. He ought,	" " 27 he will be able to

- B) 1 mark each (8 marks); no penalty for spelling errors.

1.C

- | | |
|-----------------|--------------------|
| 2a. ...name | e. author number |
| b. forename | f. title (of book) |
| c. call number | g. imprint |
| d. class number | |

- c) 1 mark for each point under each item (7 marks); no penalty for spelling errors; notes in any order.

- 3a. Name of author
 b. title (of books)
- 4c. dictionaries
 d. old editions of books
 e. any one; periodicals/professional journals
- 5f. author (name) catalogue
 6g. encyclopedia

.../

Appendix 9

Freshman English Examination (Oral Test set)

-
FRESHMAN ENGLISH EXAMINATION
(Oral Test)

Student's Name: _____
I.D.No.: _____
Faculty: _____
Department: _____
Programme (Degree/Diploma): _____
Section: _____

General Direction:

This test booklet contains 3 sections (i.e, A-C). You have 20 minutes to attempt all the tasks according to the instructions for each section with a partner. But first you have 10 minutes to silently study all the tasks before you start.

Section A

In this section the clarity and fluency of your speech is tested. Please look at your reading passage. Alem and Helina, who are ~~Students~~ at Addis Ababa University, meet and talk in the dormitory for the first time. The examiner will read Alem's part and you must read the part marked CANDIDATE. Now silently study the passage before you start reading aloud. You may write on it if you like. Remember, you will have to read the part marked CANDIDATE.

.../

The Dialogue

Alem: Hi, My name's Alem.
Candidate: Hi, I'm Helina.
Alem: Seems to me we're in the same dorm.
Candidate: That's right. Are you a freshman too?
Alem: No. I'm in the second year now.
Candidate: I see.
Alem: Well, if we're going to be dorm mates, I suppose we'll have to get on together.
Candidate: Oh, yes er would you like a cigarette?
Alem: You smoke?
Candidate: Er.... as a matter of fact, I do.
Alem: Look. Make sure you always open the window when you smoke.
Candidate: Well don't worry. I'll make sure I do that.
Alem: Um..... Do you like reading in the dormitory or in the library?
Candidate: Um I don't mind reading in the dormitory but I prefer the library.
Alem: Well, I only go to the library to look for books. I read in this room, so you'll have to be quiet.
Candidate: Ok. That's no problem.
Alem: Another thing. What time do you go to sleep in the evening?
Candidate: At about 10:30. I don't like staying up late. I'd rather get up early in the morning and read a few pages then.
Alem: That suits me. So the lights can be off before midnight.
Candidate: Yeah. That'll be fine with me.
Alem: Do you chew chat? See, I hate that thing. So if you do
Candidate: No, I don't. Besides, it's against the rules to chew chat here.....
Alem: Yeah. That's true. But.... will you have many visitors here?
Candidate: Um.... A few I guess.
Alem: Look. We'd like it quiet here. We shouldn't have more than two visitors a day.
Candidate: Er.... Yes, about two or three a day.

- 3 -

Section 2

On this card you find two lines from two different dialogues for each item. You must listen to the examiner and respond with the appropriate line by choosing from the two given. Now silently study the tasks before you start.

Example: On your card you find:

1. No. I didn't,
OR
It was good to see you too.

You hear the examiner say: "1. It was good to see you yesterday."
You must respond by saying: "It was good to see you too."
Candidate's card.

1. You mean that man up in Ambo?

OR

Why? What's up?

2. All weekend?

OR

I read about him in the paper this week. He'd been blind for about eight or nine years.

3. I'm not sure I can be gone all weekend. I promised to help my sister move Saturday.

OR

Right. He didn't have to wear a hearing aid anymore.

4. Me too!

OR

I might be able to do that. Let me check and I'll call you at home tonight.

...!

- 4 -

Section C

You are provided with part of a CLASS SCHEDULE as a source of information. Use this source to orally ask and answer questions in turns. Your partner has the information you need, and you have the information he/she needs. Double question marks(??) are placed where information is missing. Now ask about five questions to find out the information. But first silently study the source before you start.

Example

Partner 1: "What section takes instructional media?"

Partner 2: "Third year students in the Department of Foreign Languages and Literature take Instructional Media."

**ADDIS ABABA UNIVERSITY
INSTITUTE OF LANGUAGE STUDIES
CLASS SCHEDULE**

Semester I Academic Year 1992/93

Department: Foreign Languages & Literature

Course No.	Course Title	Credit	Period	Section	No. of Students	Instructor	Room
	Year FLEn III						
Educ 321	Instructional Media	3	MWF6	??	65	Birara	
Ed 351	Int. to Educational Psychology	3	MWF3	FLEn III	45	Daniel D.	
Hist 301	Eth. & the Horn 1527-1855	??	MWF3		20	Tessema T.	
Hist 321	Colonial Africa 1884-1918	3	MWF7		20	Abdussamad	
Geog 211	Cultural Geography	3	??		20	Hussein A.	
Geog 221	Introduction to Climate	3	MWF2		20	Daniel G.	
FLEn 303	Prose Fiction	3	TTHS1	FLEn III	38	Uniyal	??
FLEn 305	Skills Dev. Methodology	3	TTh6,7	FLEn III	38	Atkins	217 OCR
ELAm 303	Amharic Poetry I	3	MWF4		??	Yilma	204 OCR
ELAm ? 7	Structure of Amharic I	3	MWF8		10	Alemayehu A.	217 OCR
??	Intermediate English II	3	TTHS3		20	Mammo/ Sharma	103 OCR
FLEn 307	Advanced Speech	3	MWF4		20	George	322 OCR
FLEn 309	??	2	MW8		20	Amare A.	216 OCR
ELEn 313	Reading Skills	2	TTh4		20	Worede	219 OCR
Eng 313	English Phonetics & Phonology	3	MWF3		20	??	204 OCR
ELAm 301	Amharic Prose Fiction I	3	M.		20	Zerihun	103 OCR

FRESHMAN ENGLISH EXAMINATION

(Oral Test)

Student's Name: _____
I.D.No.: _____
Faculty: _____
Department: _____
Programme (Degree/Diploma): _____
Section: _____

General Direction:

This test booklet contains 3 sections (i.e, A-C). You have 20 minutes to attempt all the tasks according to the instructions for each section with a partner. But first you have 10 minutes to silently study all the tasks before you start.

Section A

In this section the clarity and fluency of your speech is tested. Please look at your reading passage. Alem and Helina, who are students at Addis Ababa University, meet and talk in the dormitory for the first time. The examiner will read Alem's part and you must read the part marked CANDIDATE. Now silently study the passage before you start reading aloud. You may write on it if you like. Remember, you will have to read the part marked CANDIDATE.

.../

The Dialogue

Alem: Hi, My name's Alem.
Candidate: Hi, I'm Helina.
Alem: Seems to me we're in the same dorm.
Candidate: That's right. Are you a freshman too?
Alem: No. I'm in the second year now.
Candidate: I see.
Alem: Well, if we're going to be dorm mates, I suppose we'll have to get on together.
Candidate: Oh, yes er would you like a cigarette?
Alem: You smoke?
Candidate: Er.... as a matter of fact, I do.
Alem: Look. Make sure you always open the window when you smoke.
Candidate: Well don't worry. I'll make sure I do that.
Alem: Um..... Do you like reading in the dormitory or in the library?
Candidate: Um I don't mind reading in the dormitory but I prefer the liberary.
Alem: Well, I only go to the library to look for books. I read in this room, so you'll have to be quiet.
Candidate: Ok. That's no problem.
Alem: Another thing. What time do you go to sleep in the evening?
Candidate: At about 10:30. I don't like staying up late. I'd rather get up early in the morning and read a few pages then.
Alem: That suits me. So the lights can be off before midnight.
Candidate: Yeah. That'll be fine with me.
Alem: Do you chew chat? See, I hate that thing. So if you do
Candidate: No, I don't. Besides, it's against the rules to chew chat here.....
Alem: Yeah. That's true. But.... will you have many vistors here?
Candidate_ Um.... A few I guess.
Alem: Look. We'd like it quiet here. We shouldn't have more than two visitors a day.
Candidate: Er.... Yes, about two or three a day.

- 3 -

Section 2

On this card you find two lines from two different dialogues for each item. You must listen to the examiner and respond with the appropriate line by choosing from the two given. Now silently study the tasks before you start.

Example: On your card you find:

1. No. I didn't.
OR
It was good to see you too.

You hear the examiner say: "1. It was good to see you yesterday."
You must respond by saying: "It was good to see you too."

Candidate's card.

1. You mean that man up in Ambo?

OR

Why? What's up?

2. All weekend?

OR

I read about him in the paper this week. He'd been blind for about eight or nine years.

3. I'm not sure I can be gone all weekend. I promised to help my sister move Saturday.

OR

Right. He didn't have to wear a hearing aid anymore.

4. Me too!

OR

I might be able to do that. Let me check and I'll call you at home tonight.

...!

- 4 -

Section C

You are provided with part of a CLASS SCHEDULE as a source of information. Use this source to orally ask and answer questions in turns. Your partner has the information you need, and you have the information he/she needs. Double question marks (??) are placed where information is missing. Now ask about five questions to find out the information. But first silently study the source before you start.

Example

Partner 1: "What section takes Skills Development Methodology,"

Partner 2: "Third year students in the Department of Foreign Languages and Literature take Skills Development Methodology."

.../

ADDIS ABABA UNIVERSITY
INSTITUTE OF LANGUAGE STUDIES
CLASS SCHEDULE

Semester I Academic Year 1992/93

Department : Foreign Languages & Literature

Course No.	Course Title	Credit	Period	Section	No. of Students	Instructor	Room
	Year FLEr III						
Educ 321	Instructional Media	3	MWF6	FLEn III	65	Birara	
Epsy 351	Int. to Educational Psychologh	3	??	FLEn III	45	Daniel D.	
Hist 301	Eth. & the Horn 1527-1855	3	MWF3		20	Tessema T.	
Hist 321	Colonial Africa 1884-1918	??	MWF7		20	Abdussamad	
Geog 211	Cultural Geography	3	TTh32		20	Hussein A.	
Geog 221	Introduction to Climate	3	MWF2		??	Daniel G.	
FLEn 303	Prose Fiction	3	TThs 1	FLEn III	38	Uniyal	322 OCR
FLEn 305	Skills Dev. Methodology	3	TTh 6,7	??	38	Atkins	217 OCR
ELAM 303	Amharic Poetry I	3	MWF 4		10	Yilme	204 OCR
??	Structure of Amharic I	3	MWF8		10	Alemayehu A.	217 OCR
FLEn 301	Intermediage English II	3	TThs 3		20	Mammo/ Shayma	103 OCR
FLEn 307	Advanced Speech	3	M WF4		20	??	322 OCR
FLEn 309	Public Relations	2	MW8		20	Amare A.	216 OCR
FLEn 313	??	2	TTh4		20	Worede	219 OCR
Ling. 313	English Phonetics & Phonology	3	MWF 3		20	Baye	204 OCR
ELAm 301	Amharic Prose Fiction I	3	MWF8		20	Zerihun	??

Appendix 10

Examiners' set for the Oral Test: source and scoring sheet,
all with instructions and evaluation guidelines

Section A

*In this section the clarity and fluency of your
speech is tested. Please look at your reading passage.
Alem and Helina, who are students at Addis Ababa
University, meet and talk in the dormitory for the first
time. I will read Alem's part and you must read the part
marked CANDIDATE. Remember, you have to read the part
marked CANDIDATE. Ready?*

The Dialogue

Alem: Hi, My name's Alem.

Candidate: Hi, I'm Helina.

Alem: Seems to me we're in the same dorm.

Candidate: That's right. Are you a freshman too?

Alem: No. I'm in the second year now.

Candidate: I see.

Alem: Well, if we're going to be dorm mates, I suppose we'll have to get on together.

Candidate: Oh, yes er would you like a cigarette?

Alem: You smoke?

Candidate: Er.... as a matter of fact, I do.

Alem: Look. Make sure you always open the window when you smoke.

Candidate: Well don't worry. I'll make sure I do that.

Alem: Um..... Do you like reading in the dormitory or in the library?

Candidate: Um I don't mind reading in the dormitory but I prefer the library.

Alem: Well, I only go to the library to look for books. I read in this room, so you'll have to be quiet.

Candidate: Ok. That's no problem.

Alem: Another thing. What time do you go to sleep in the evening?

Candidate: At about 10:30. I don't like staying up late. I'd rather get up early in the morning and read a few pages then.

Alem: That suits me. So the lights can be off before midnight.

Candidate: Yeah. That'll be fine with me.

Alem: Do you chew chat? See, I hate that thing. So if you do

Candidate: No, I don't. Besides, it's against the rules to chew chat here.....

Alem: Yeah. That's true. But.... will you have many visitors here?

Candidate: Um.... A few I guess.

Alem: Look. We'd like it quiet here. We shouldn't have more than two visitors a day.

Candidate: Er.... Yes, about two or three a day.

Section D

On your card you find a double dialogue, ie, the card has the lines from two different dialogues. Therefore, you must listen to me and respond with the appropriate line by choosing from the two lines given. Ready?

Examiner's card and Answer Key

1. Hi. I'm glad I ran into you.

Response: Why? What's up?

OR

Did you hear about that guy who was struck by lightning?

Response: You mean that man up in Ambo?

2. How'd you like to go on a trip this weekend?

Response: All weekend?

OR

That's the one; the blind guy who could see again.

Response: I read about him in the paper this week. He'd been blind for about eight or nine years.

3. Well, just Saturday and Sunday.

Response: I'm not sure I can be gone all weekend. I promised to help my sister move Saturday.

OR

Uh-huh. Wasn't he also able to hear again after the lightning hit him?

Response: Right. He didn't have to wear a hearing aid anymore.

4. How about Sunday, then? We could start early in the morning.

Response: I might be able to do that. Let me check and I'll call you at home tonight.

OR

I think it was a miracle!

Response: Me too!

5. Great! I'll talk to you later. --- That's it!

Or

Yes! ... That's it!

Section C

You are provided with part of a CLASS SCHEDULE as a source of information. Use this source to orally ask and answer questions in turns. Your partner has the information you need, and you have the information he/she needs. Double question marks (??) are placed where information is missing. Now ask about five questions to find out the information. Ready? Who would like to start asking?

Note: The examiner gets the candidate rehearse the examples. That is, partner one raises the question at asterisk one (*₁), and partner two raises another question at asterisk two (*₂), both in the column for "section". They take turns to ask and answer questions.

.../

ADDIS ABABA UNIVERSITY
INSTITUTE OF LANGUAGE STUDIES

Class Schedule

Semester I Academic Year 1992/93

Department: Foreign Languages & Literature.

Course No.	Course Title	Credit	Period	Section	No. of Students	Instructor	Room
	Year FLEn III						
Educ 321	Instructional Media	3	MWF6	FLEn III ¹	65	Birara	
EPSY 351	Int. to Educational Psy.	3	MWF3*2	FLEn III	45	Daniel D.	
Hist 301	Eth. & the Horn 1527-1855	3*1	MWF3		20	Tessema T.	
Hist 321	Colonial Africa 1884-1918	3*2	MWF7		20	Abdussamad	
Geog 211	Cultural Geography	3	T ThS2*1		20	Hussein A.	
Geog 221	Introduction to climate	3	MWF2		20*2	Daniel G.	
FLEn 303	Prose Fiction	3	TTHS 1	FLEn III	38	Uniyal	322 OCR*1
FLEn 305	Skills Dev. Methodology	3	TTH 6,7	FLEn III*2	38	Atkins	217 "
ELAm 303	Amharic Poetry I	3	MWF 4		10*1	Yilma	204 "
ELAm 307*2	Structure of Amharic I	3	MWF 8		10	Alemayehu A.	217 "
FLEn 301*1	Intermediate English II	3	TThS 3		20	Mamo/Sharma	103 "
FLEn 307	Advanced Speech	3	MWF 4		20	Georg*2	322 "
FLEn 309	Public Relations*1	2	MW 8		20	Amaré A	216 "
FLEn 313	Reading Skills *2	2	TTh 4		20	Worele	219 "
Ling 313	English Phonetics & Phon.	3	MWF 3		20	Baye*1	204 "
ELAm 301	Amharic Prose Fiction I	3	MWF 8		20	Zerihun	103 "*2

- 1 -

Evaluation Sheet

*Please find the evaluation guidelines on page 2.

Student's Name and I.D.No.	Section A	\bar{X}	Section B	\bar{X}	Section C	\bar{X}	\bar{X} s
1. _____	()	()	()	()	()	()	()
2. _____	()	()	()	()	()	()	()
3. _____	()	()	()	()	()	()	()
4. _____	()	()	()	()	()	()	()
5. _____	()	()	()	()	()	()	()
6. _____	()	()	()	()	()	()	()
7. _____	()	()	()	()	()	()	()
8. _____	()	()	()	()	()	()	()
9. _____	()	()	()	()	()	()	()
10. _____	()	()	()	()	()	()	()
11. _____	()	()	()	()	()	()	()
12. _____	()	()	()	()	()	()	()
13. _____	()	()	()	()	()	()	()
14. _____	()	()	()	()	()	()	()
15. _____	()	()	()	()	()	()	()
16. _____	()	()	()	()	()	()	()
17. _____	()	()	()	()	()	()	()
18. _____	()	()	()	()	()	()	()
19. _____	()	()	()	()	()	()	()
20. _____	()	()	()	()	()	()	()
21. _____	()	()	()	()	()	()	()
22. _____	()	()	()	()	()	()	()
23. _____	()	()	()	()	()	()	()
24. _____	()	()	()	()	()	()	()
25. _____	()	()	()	()	()	()	()
26. _____	()	()	()	()	()	()	()
27. _____	()	()	()	()	()	()	()
28. _____	()	()	()	()	()	()	()
29. _____	()	()	()	()	()	()	()
30. _____	()	()	()	()	()	()	()

- 2 -

Evaluation Guidelines

- * Score each section (A,B,C) on a 5 point scale
- * Work out the average score of the 3 judges for each section
- * Write the average score for each section

Section A (Reading aloud a dialogue)

- 5= Very comprehensible; very clear and fluent
- 4=Reasonably comprehensible; good clarity and fluency
- 3=Comprehensible; fairly clear and fluent
- 2=Partially comprehensible; poor in clarity and fluency
- 1/0= Incomprehensible; very poor in clarity and fluency

Section B (Responding to double dialogue)

- 5=Very appropriate (4 correct)
- 4=Reasonably appropriate (3 ")
- 3=Appropriate (2 ")
- 2= Hardly appropriate (1 ")
- 1/0= Inappropriate (none ")

Section C (Asking and answering questions from source)

- 5= Understands data, and is able to ask and answer questions very well
- 4=Understands data, and is able to ask and answer questions reasonably well
- 3=Understands data, and is able to ask and answer questions
- 2=Understands data partially, and is rather poor in asking and answering questions
- 1/0=Hardly understands data, and is very poor in asking and answering questions

Appendix 11
External written test (WT2)

GENERAL DIRECTION:

This test booklet contains two sections with four parts (ie, Section 2, Parts A and B; Section 3, Parts A and B). You have one hour and ten minutes to attempt all the questions according to the instructions for each section and/or part.

- 6 -

SECTION 2

STRUCTURE AND WRITTEN EXPRESSION

Time- 25 minutes

This section tests your ability to identify appropriate forms of formal written English. There are two parts to this section.

2: Part A

Directions: Questions 1-15 contain sentences that are incomplete in some way. You choose the ONE answer that completes each sentence properly.

Example I

The horizon appears to be curved when viewed -----a high place.

- (A) with
- (B) from
- (C) on
- (D) out of

The sentence should read, "The horizon appears to be curved when viewed from a high place." Therefore, (B) is the correct answer.

Example II

-----the meaning of vocabulary from context is an important skill.

- (A) Determining
- (B) Having determined
- (C) It is determined
- (D) Determination.

The sentence should read, "Determining the meaning of vocabulary from context is an important skill." Therefore, (A) is the correct answer.

Now begin work on the questions.

1. The Canterbury Tales, -----
by English poet Geoffrey Chaucer, are a fine example of Middle English.
 - (A) a story set of
 - (B) as sets of stories
 - (C) a set of stories
 - (D) stories in a set

2. Albert Camus, a French journalist who won the 1957 Nobel prize
for literature, -----about individual freedoms and
alienation from society.
 - (A) and wrote passionately
 - (B) writing passionately
 - (C) he wrote passionately
 - (D) wrote passionately

- 9 -

3. Cardiac arrest occurs _____ heart muscles stop beating during a heart attack or after a severe accident.
- (A) when
 - (B) in that
 - (C) although
 - (D) due to
4. Harvard University is one of the oldest _____ private educational institutions in the United States.
- (A) but rich
 - (B) and richest
 - (C) yet richer
 - (D) and the rich
5. It is not clear _____ the form of humor known as a limerick started.
- (A) if
 - (B) how
 - (C) which
 - (D) whenever
6. _____ only U.S. state that is not situated on the mainland of North America.
- (A) Hawaii, the
 - (B) It is Hawaii the
 - (C) Hawaii is the
 - (D) Because Hawaii is the
7. Hannibal, the military genius of Carthage, was best known _____ the mountains of the Pyrennees and Alps with elephants.
- (A) as crossed
 - (B) for crossing
 - (C) with the crossing of
 - (D) having crossed
8. Of all types of health maintenance organizations, _____.
- (A) the most common being a prepaid group practice plan
 - (B) a prepaid group practice plan was commonly
 - (C) there is the most common a prepaid group practice plan
 - (D) a prepaid group practice plan is the most common
9. The strong beam of light from a lighthouse is used by sailors _____.
- (A) to determining their location
 - (B) for determination their location
 - (C) in determining their location
 - (D) while determined their location

- 10 -

10. -----travels 5.8 trillion miles in one year has been scientifically proven.
- (A) That light
 - (B) Light
 - (C) For light to
 - (D) When light
11. A caricature is a form of art involving a picture ----- the physical nature of an important person.
- (A) exaggerated
 - (B) that exaggerates
 - (C) having exaggerated
 - (D) exaggerating that
12. Resources such as factories or equipment used to produce goods and services are -----as capital.
- (A) referred
 - (B) referring in
 - (C) referred to
 - (D) being referred
13. In the 1830s railroads ----- canals as the most important means of transportation in the United States.
- (A) replacing by
 - (B) were replaced
 - (C) was replacing
 - (D) replaced
14. The human body acquires nutrients from food particles -----the blood.
- (A) carrying
 - (B) carried in
 - (C) carrying by
 - (D) has carried
15. Rarely ----- primitive tribes of Africa, Asia, and the Pacific Islands practice cannibalism today.
- (A) have
 - (B) are
 - (C) do
 - (D) is it

- 11 -

2: Part B

Directions: Questions 16-40 contain sentences that are incorrect in some way. You choose the one underlined part that is wrong.

Example I

Particles of salt dissolved in water retains their salinity.
 A B C D

The sentence should read, "Particles of salt dissolved in water retain their salinity." Therefore, (C) is the proper choice.

Example II

The emissions from a diesel-powered automobile are not as dangerous
 A B
 as that from a gasoline-powered vehicle.
 C D

The sentence should read, "The emissions from a diesel-powered automobile are not as dangerous as those from a gasoline-powered vehicle." Therefore, (C) is the proper choice.

Now begin work on the questions.

16. The Persian poet Hafiz wrote a popular collecting of 700 poems
 A B C
 called a devan,
 D
17. All the necessary ingredient for the survival of plants and
 A B C
 animals are found within a habitat.
 D
18. The first person to sign Declaration of Independence in 1776 was John
 A B C D
 Hancock.
19. The heart is a large hollow muscle that pumps blood out the body.
 A B C D
20. Hares resemble rabbits yet having longer and larger legs and ears
 A B C
 than their cousins.
 D
21. The field of linguistic concerns itself with the analysis and use
 A B C D
 of language.
22. A migraine, considered to be the bad kind of headache, can cause
 A B
nausea in the most extreme cases.
 C D

- 12 -

23. A carnival is a form of amusement involving games, rides, shows, and exhibiting in an outdoor area.
A B C D
24. Carbohydrates, which include all sugars and starches, are the source main of energy for the body.
A B C D
25. The alphabet invented about 1500 B.C. by Semitic people inhabiting the Middle East.
A B C D
26. The importance of historical perspective in the understanding of human culture was a primary concern of the well-known German philosophy George Hegel.
A B C D
27. Advances in computer graphics have brought on a new era in the field of cartoon animated.
A B C D
28. Since the early 1900s Luther Burbank bred a spineless variety of cactus that can be eaten as food.
A B C D
29. Manuscript writing consists of letters written straight up and down and not connecting to each other.
A B C D
30. Calamity Jane was a famous American frontierswoman who was adept with handling a horse and rifle.
A B C D
31. Calculus, one of the most important mathematical areas, are used for measuring changing quantities of speed and distance.
A B C D
32. Lacquer is a shiny substance that protects the surface of metals, woods, and porcelain from wear or tear.
A B C D
33. In large amounts, caffeine can induce nervousness and caused loss of sleep.
A B C D
34. Latin America comprises of that part of the Western Hemisphere south of the United States.
A B C D
35. Latium, the area of ancient Italy after which the Latin language was named, it was part of the Roman Empire.
A B C D

- 13 -

36. Leather is the durable and long-lasting material obtained from
A B C
the skin of animals.
D
37. Contrary to popular belief, the hump of a camel contains fatty
A B
tissue only and has no water inside of them.
C D
38. The domed cathedral Hagia Sophia in Constantinople is a fine
A B
example of the Byzantine style of architectural.
C D
39. The process of replacing lost parts of bodies or doing new forms
A B C
is called regeneration.
D
40. In practical all societies, females live longer than their male
A B C D
counterparts.
D

- 14 -

SECTION 3
VOCABULARY AND READING COMPREHENSION

Time - 45 minutes

This section tests your understanding of written English. It is divided into two parts.

. 3: Part A

Directions: In questions 1-30 each sentence has a word or phrase underlined. Below each sentence are four choices, and you must select the one that MOST CLOSELY MEANS THE SAME AS the underlined word.

Example

Meteorologists use balloons to forecast weather patterns.

- (A) cause
- (B) predict
- (C) create
- (D) test

(B) is the correct answer because the word "predict" most closely means "forecast."

Now begin work on the questions.

1. The Code of Hammurabi in ancient Babylonia had great influence on the civilizations of nearly all Near Eastern countries.
 - (A) clearly
 - (B) quite
 - (C) closely
 - (D) almost
2. One-third of the Earth's surface area of approximately 197 million square miles is devoted to farming.
 - (A) totally
 - (B) largely
 - (C) exactly
 - (D) roughly
3. One of the oldest forms of folk art is handicraft, the creation of objects with the hands.
 - (A) invention
 - (B) activity
 - (C) hobby
 - (D) elevation
4. Indication that cannibalism was once a religious practice dates back almost 500,000 years.
 - (A) custom
 - (B) assignment
 - (C) recital
 - (D) business

- 15 -

5. Coffee contains the odorless, bitter substance caffeine, which acts as a stimulant to the human heart and nervous system.
- (A) sweetened
 - (B) acrid
 - (C) delicious
 - (D) pure
6. Electrical conductors such as copper and aluminum need to be insulated from the possible signal interference of other conductors.
- (A) connected
 - (B) realigned
 - (C) protected
 - (D) redistributed
7. Alex Haley spent 12 years investigating his family history from Africa in the 1700s to the present day.
- (A) developing
 - (B) researching
 - (C) categorizing
 - (D) declaring
8. An heir is defined as the legal beneficiary of the money or property of a person who has died without leaving a will.
- (A) relative
 - (B) associate
 - (C) recipient
 - (D) ancestor
9. John Gower's Confessio Amantis is an exemplary poem combining Biblical, medieval, and mythological stories to explore the problems of romantic love.
- (A) a detailed
 - (B) a model
 - (C) a resolute
 - (D) a simple
10. The two basic principles of aerodynamics, lift and drag, are applied to the design of all airplanes, vehicles, and buildings.
- (A) registered in
 - (B) combined with
 - (C) relegated to
 - (D) employed in
11. The poetry of T.S. Eliot was known for its complicated style and extensive use of symbolism.
- (A) widespread
 - (B) accurate
 - (C) limited
 - (D) lengthened

- 16 -

12. The sharp bristles of a cactus prevent desert animals in search of moisture from devouring it.
- (A) destroying
 - (B) consuming
 - (C) cultivating
 - (D) manipulating
13. The theory of laissez-faire economics holds that government should not interfere with competition in an open market.
- (A) a prescribed
 - (B) a free
 - (C) a spacious
 - (D) a balanced
14. A wide range of conditions, including injury, infection, and allergic reaction, can bring on a headache.
- (A) induce
 - (B) alleviate
 - (C) negate
 - (D) foment
15. The inherent freedom of economic choice within the capitalist system promotes competition in the production of goods and services.
- (A) chronic
 - (B) regulated
 - (C) intrinsic
 - (D) novel
16. Robert Heinlein, a well-known American science fiction writer, made his stories plausible by explaining the nature of future societies in great detail.
- (A) credible
 - (B) entertaining
 - (C) unrealistic
 - (D) zealous
17. Latex, the milklike juice given off by plants and trees of the sapodilla family, is used in producing gums and rubbers.
- (A) covered
 - (B) emitted
 - (C) attached
 - (D) delivered
18. The breakdown of the ozone layer surrounding the Earth correlates with the increased occurrence of skin cancer.
- (A) movement
 - (B) deterioration
 - (C) failure
 - (D) itemization

- 17 -

19. The Mississippi River has always played an indispensable role in trade and commerce through the nation's heartland.
- (A) a developmental
 - (B) a thorough
 - (C) an elevated
 - (D) an essential
20. The action of water vapor and carbon dioxide in keeping heat created by sunlight from escaping back into space is appropriately named the greenhouse effect.
- (A) incorrectly
 - (B) usually
 - (C) deliberately
 - (D) suitably
21. Halloween customs developed from ancient religious festivals that occurred the day before All Saints' Day.
- (A) celebrations
 - (B) services
 - (C) documents
 - (D) traditions
22. The work of Frank Lloyd Wright, one of America's greatest architects, has enduring beauty for those who appreciate his unique style.
- (A) captive
 - (B) lasting
 - (C) regressing
 - (D) serene
23. Many lawsuits are resolved through settlement prior to court due to the effort and expense involved in trial.
- (A) debate
 - (B) amelioration
 - (C) dismissal
 - (D) compromise
24. The writ of habeas corpus is a basic personal freedom under American law that compels police authorities to justify incarcerating an individual.
- (A) releasing
 - (B) suspecting
 - (C) restraining
 - (D) imprisoning
25. The time between submission of a manuscript to a publisher and eventual publication can sometimes take several years.
- (A) a prologue
 - (B) a composition
 - (C) an addendum
 - (D) an essay

- 18 -

26. Helen Hayes, one of America's outstanding actresses, won an Academy Award for the first film performance of her long and distinguished career.
- (A) preeminent
 - (B) charming
 - (C) reclusive
 - (D) profound
27. Cursive writing involves letters angled to one side or another and connected together.
- (A) supported
 - (B) regulated
 - (C) slanted
 - (D) relegated
28. Prehistoric man drew simplified representations called pictographs on cave walls to depict common articles and ideas.
- (A) outlines
 - (B) images
 - (C) coordinates
 - (D) illusions
29. Northern carpetbaggers were viewed by Southerners as opportunistic intruders seeking wealth and political power in the wake of the Civil War.
- (A) generous
 - (B) self-serving
 - (C) invading
 - (D) decisive
30. Hairdressing, the art of caring for the hair, has existed as a bona fide profession since the 1700s.
- (A) dominant
 - (B) dubious
 - (C) casual
 - (D) genuine

- 19 -

3: Part B

Directions: In this part you will read passages on various topics and answer questions about them. Some answers to the questions are directly stated, while others are only suggested.

Read the following passage:

The pituitary gland, one of the body's key organs, generates a number of hormones that help control body function. Sometimes known as the hypophysis, the gland has two main parts—the anterior lobe and the posterior lobe. It is about the size of a pea and is located near the center of the skull.

Example I

According to the passage, what is another name for the pituitary gland?

- (A) The anterior lobe
- (B) The hormone
- (C) The hypophysis
- (D) The posterior lobe

The passage states that the pituitary gland is also known as the hypophysis. Therefore, the correct answer is (C).

Example II

It can be inferred from the passage that the pituitary gland is

- (A) quite small
- (B) divided into several parts
- (C) very large
- (D) composed of a number of hormones

The passage states that the pituitary gland is about the size of a pea, which suggests that it is quite small. The correct answer is (A).

Now begin work on the questions.

- 20 -

Questions 31-35

(5) The United States population is growing older. In 1987, 12 percent of Americans were 65 years old or older, compared with 8 percent in 1950. Population experts at the U.S. Bureau of the Census expect this percentage to continue to rise gradually, reaching 14 percent in 2010, then to skyrocket during the next 20 years, reaching 21 percent by 2030. This "graying of America" has generated concerns about whether the best really is yet to be, about how well off tomorrow's elderly will be. There also are questions about the impact of an aging population on the rest of society.

(10) The graying of America has two causes. First, advances in medical care have enabled people to live longer. In the United States in 1900, the average life expectancy at birth was 47.3 years. By 1985, the latest year for which figures are available, it has climbed to 74.7 years.

(15) Second, the U.S. birthrate rose in the mid-1900s, interrupting a long, slow decline. A dramatic increase occurred during this brief rise. From 1945 to 1947, the rate jumped from 20.4 births for every 1,000 people to 26.6. In 1957, the birthrate was still high - 25.3-but then began to decline. The birthrate dropped to the 1933 level of 18.4 in 1966, and by 1985, was 15.7. Most experts see little change in the downward trend of the birthrate in the foreseeable future.

31. What is the author's main purpose in the passage?
- (A) To bring attention to the needs of the elderly
 (B) To contrast birthrates and life expectancy rates
 (C) To explain how and why the American population is aging
 (D) To show the percentages of age groups in the general population
32. During which period of time will the percentage of Americans 65 years old or older increase the most?
- (A) 1970-1990
 (B) 1990-2010
 (C) 2010-2030
 (D) 2030-2050
33. In the first paragraph, the phrase "the graying of America" refers to
- (A) the uncertainty of the future for older Americans
 (B) the effects of population on the health of the elderly
 (C) statistics released by the U.S. Bureau of the Census
 (D) the increasing number of elderly people in the population

- 21 -

34. About how long did it take the birthrate to return to the 1933 level?
- (A) 20 years
 - (B) 30 years
 - (C) 40 years
 - (D) 50 years
35. Which of the following conclusions about the birthrate is supported by the passage?
- (A) The long-range decline in the birthrate will continue in the next century.
 - (B) Another dramatic increase in the birthrate will occur before the 21st century.
 - (C) The birthrate will continue to vary greatly.
 - (D) An increase in the birthrate depends mainly on advances in medical care.

- 22 -

Questions 36-41

- In countries having democratic forms of government, political parties compete in open elections for the right to run the government. The primary functions of parties include nominating suitable candidates for public office, selecting issues for public debate and persuading persons who have registered to vote to elect its slate. Party leaders are also responsible for raising money to finance costly political campaigns.
- (5)

- Once elected, party officials try to implement the policies of their party in the operation of government. If the party has just taken over control of the government, smooth operation of the system and implementation of policy might take months to achieve. In the case of a party that retains its hold on the government through reelection, the clear mandate of the people normally results in a flurry of new laws that can influence the social and economic climate of a country for years.
- (10)

- The primary function of the party or parties not elected is to question the policies of the party in control. The minority parties publicize what they consider deficiencies in the majority party's program -and offer voters an alternative. This healthy opposition serves as a check against the excesses of power seen in one-party systems.
- (15)

36. Which of the following is the best title for the passage?
- (A) Democratic Forms of Government
 (B) The Roles of Political Parties
 (C) Election of Government Officials
 (D) Functions of Political Campaigns
37. According to the passage, elections in a democracy are
- (A) run by party officials
 (B) organized by potential candidates
 (C) vulnerable to the excesses of power
 (D) open to individuals who have signed up to vote
38. Which of the following is NOT mentioned as one of the functions of political parties?
- (A) Raising funds for financing government elections
 (B) Choosing candidates to run for office
 (C) Deciding on issues for public discussion
 (D) Convincing people to vote for its slate of candidates
39. According to the passage, a party reelected to power
- (A) may require months to recover from the effects of a difficult election contest
 (B) can have great impact on the future health and welfare of a nation
 (C) attacks the deficiencies of the minority party's program
 (D) often suffers the excesses in power of one-party systems

- 23 -

40. It can be inferred from the passage that a democratic country
- (A) endures none of the disadvantages of a one-party system
 - (B) must constantly regulate the activities of its political parties
 - (C) maintains at least two political parties acting in opposition
 - (D) benefits from the fund-raising activities of political parties
41. According to the passage, what is the role of a political party NOT in control of the government?
- (A) It serves as an alternative to the ruling party and its policies.
 - (B) It helps to implement governmental policy until the next election.
 - (C) It questions the legality of the ruling party's actions.
 - (D) It follows the people's mandate by legislating new laws.

Questions 42-48

- The phenomenon of lightning has been a source of danger and mystery through the centuries, and only recently have scientists begun to understand its true nature. We know now that lightning is a huge electrical spark caused by the interaction of electrically charged particles in the atmosphere. During a thunderstorm, heavy particles take on a negative charge and fall to the bottom of clouds, while particles holding a positive charge rise to the top. Lightning results when energy flows between the two types of charges.
- (5)

- What is still not understood is the process by which particles become electrically charged. Most scientists believe that a cloud's light, rising moisture and tiny pieces of ice collide with hail and other heavy, falling particles. The collision may cause these elements to take on either a positive or negative charge.
- (10)

- The most feared form of lightning is that which strikes the ground in one or more electrical discharges called strokes. Any one of these strokes is capable of causing death, destroying property, or igniting fires. It is interesting to note that the bright flash of light produced by a stroke actually occurs as the discharge returns to the sky and not as it descends towards the ground. This return stroke heats surrounding air and causes it to expand, thus producing a wave of pressure called thunder.
- (15)
- (20)

42. Which of the following is the best title for the passage?
- (A) A Dangerous Phenomena
(B) Atmospheric Occurrences
(C) The Nature of Lightning
(D) Scientific Mysteries
43. The passage states that lightning is
- (A) a recent source of mystery
(B) a giant spark of electricity
(C) an electrically charged particle
(D) a type of thunderstorm
44. According to the passage, lightning results from
- (A) a flow of energy between positive and negative charges
(B) discharges in a heavy thunderstorm
(C) the rising and falling of heavy particles
(D) great amounts of electricity in the atmosphere
45. It can be inferred that positively charged particles are
- (A) the primary cause of lightning
(B) found only during thunderstorms
(C) not involved in the production of lightning
(D) lighter than negatively charged ones

- 25 -

46. According to the passage, particles become electrically charged when
- (A) light, ascending matter collides with heavy, falling elements
 - (B) energy flows between light and heavy charged particles
 - (C) heavy particles rise to the tops of clouds
 - (D) a return stroke heats the surrounding air and causes it to expand
47. The paragraph following this passage most likely discusses
- (A) the nature of thunder
 - (B) other types of lightning
 - (C) the dangerous effects of electrical discharges
 - (D) another natural phenomenon
48. The author implies that most people believe they observe lightning
- (A) as being slower than the speed of light
 - (B) only in conjunction with the sound of thunder
 - (C) as it rushes down to strike the earth
 - (D) in the form of strokes released in the atmosphere

- 26 -

Questions 49-53

- Abraham Lincoln, the 16th president of the United States, was one of the truly great men of all time. He preserved the American Union during the Civil War and proved to the world that democracy can be a lasting form of government. Lincoln's
- (5) Gettysburg Address, and many of his other speeches and writings, are classic statements of democratic beliefs and goals.

- Lincoln's outstanding asset was insight. Lincoln realized at the beginning of the Civil War that the Union must be saved. The United States was the only important democracy in the world.
- (10) Lincoln knew that self-government would be proved a failure if the nation could be destroyed by a minority of its own people. He determined that the nation, and democracy, would not be destroyed.

- Lincoln's second greatest asset was his ability to express his convictions so clearly, and with such force, that millions of Americans made them their own. Lincoln would have been surprised that some of his speeches came to be honored as great literature. He sought only to be understood, and to convince.
- (15)

- Lincoln's third great source of strength was his iron will. The Civil War had to be carried on until the Union was restored. At times, people in the North wavered in this purpose. Lincoln never doubted that in the end, right would make might, and the North would triumph. His unyielding faith in victory helped to win victory.
- (20)

49. According to the passage, what was the Gettysburg Address?
- (A) A declaration of war against the South
 (B) A famous talk by Lincoln on democratic principles
 (C) One of the essays on government written by Lincoln
 (D) Lincoln's speech at his inauguration
50. According to the passage, what was Abraham Lincoln's greatest strength?
- (A) The ability to see the true nature of a situation
 (B) The capacity to express his views in public
 (C) His success in becoming president of the United States
 (D) The impact of his speeches and writings
51. The author probably feels that Lincoln
- (A) was at times too stubborn for his own good
 (B) never fully realized the importance of self-government
 (C) was as great a writer as he was a politician
 (D) was single-handedly responsible for saving the Union
52. The passage would most likely be found in a textbook on which of the following subjects?
- (A) Historical literature
 (B) Public administration
 (C) The Civil War
 (D) Political science

- 27 -

53. According to the passage, Abraham Lincoln believed that the North
- (A) would quickly triumph over the minority of people trying to destroy the Union
 - (B) would achieve victory through the moral power of its cause
 - (C) did not have the strength to continue the Civil War for a long time
 - (D) could waver in its purpose and lose the struggle

- 28 -

Questions 54-60

Radiocarbon is a radioactive isotope with an atomic weight of 14, which makes it heavier than ordinary carbon. Radiocarbon forms when cosmic rays, or high-energy atomic particles, collide with the Earth's atmosphere. This collision causes atoms to disintegrate into smaller elements. One of these elements, the neutron, smashes into the nuclei of nitrogen atoms and, in the process of being absorbed into the nuclei, causes a proton element to be released. In this manner a nitrogen atom turns into a radiocarbon atom.

(5)

Radiocarbon is found in all living matter. For every trillion molecules of carbon dioxide gas, the atmosphere contains about one radiocarbon atom. Plants assimilate radiocarbon from carbon dioxide in the air, and humans absorb it mainly from food made from plants.

(10)

Radiocarbon is very useful in establishing the age of old objects. The technique of radiocarbon dating was developed by an American chemist, William F. Libby, in the late 1940s. He discovered that radioactive carbon atoms decay at a regular rate over long periods of time. After about 5,700 years, half the radiocarbon in dead material disappears. After 11,400 years, half the remaining material is gone. Using his method, archeologists have been able to determine the age of objects up to 50,000 years old.

(15)

54. The main topic of this passage is
- (A) the weight of radioactive isotopes
 - (B) differences between radioactive carbon and regular carbon
 - (C) the origin and uses of radiocarbon isotopes
 - (D) forms of cosmic rays
55. According to the passage, what happens when atomic particles strike the Earth's outer layer?
- (A) Atoms break down into smaller components.
 - (B) Radiocarbon becomes heavier.
 - (C) They are absorbed by all living matter.
 - (D) Protons are released into the atmosphere
56. According to the passage, radiocarbon results when
- (A) it is heavier than regular carbon
 - (B) atomic particles escape the Earth's atmosphere
 - (C) a neutron hits the nucleus of a carbon atom
 - (D) a nitrogen atom absorbs a neutron and releases a proton
57. According to the passage, how do people primarily take in radiocarbon?
- (A) From animal food
 - (B) By breathing the air
 - (C) From carbon dioxide molecules
 - (D) From edible vegetation

- 29 -

58. Compared to carbon dioxide gas, radiocarbon
- (A) is a rare element
 - (B) is found in equal quantities
 - (C) has unusual properties
 - (D) is a very common element
59. In line 11, the word "assimilate" could best be replaced by which of the following?
- (A) produce
 - (B) transfer
 - (C) reduce
 - (D) absorb
60. It can be inferred from the passage that an object that is 11,400 years old
- (A) contains half its original radiocarbon content
 - (B) has no radiocarbon left
 - (C) has about a quarter of its radiocarbon remaining
 - (D) has less than an eighth of its initial radiocarbon matter

Appendix 12

Answer sheet for external written test

Page 1

ANSWER SHEET

Student's Name; _____
 I.D. No. : _____
 Faculty: _____
 Department: _____
 Programme (Degree/Diploma): _____
 Section: _____

General Direction: Please mark your answer for each item by putting a cross (X) on the letter of your choice. Use pencil.

Example: A C D; that is, your answer is "B"

Section 2: Structure and written Expression.

PART A

- | | | | |
|-------------|-------------|-------------|-------------|
| 1. A D C D | 2. A B C D | 3. A E C D | 4. A B C D |
| 5. A B C D | 6. A E C D | 7. A B C D | 8. A B C D |
| 9. A B C D | 10. A B C D | 11. A B C D | 12. A E C D |
| 13. A B C D | 14. A B C D | 15. A E C D | |

PART B

- | | | | |
|-------------|-------------|-------------|-------------|
| 16. A E C D | 17. A E C D | 18. A B C D | 19. A B C D |
| 20. A B C D | 21. A B C D | 22. A B C D | 23. A E C D |
| 24. A E C D | 25. A E C D | 26. A E B D | 27. A B C D |
| 28. A B C D | 29. A E C D | 30. A B C D | 31. A B C D |
| 32. A B C D | 33. A B C D | 34. A B C D | 35. A E C D |
| 36. A B C D | 37. A E C D | 38. A E C D | 39. A E C D |
| 40. A B C D | | | |

SECTION 3: Vocabulary and Reading Comprehension

PART A

- | | | | |
|-------------|-------------|-------------|-------------|
| 1. A B C D | 2. A B C D | 3. A B C D | 4. A E C D |
| 5. A B C D | 6. A E C D | 7. A B C D | 8. A B C D |
| 9. A B C D | 10. A B C D | 11. A B C D | 12. A E C D |
| 13. A B C D | 14. A E C D | 15. A E C D | 16. A E C D |
| 17. A E C D | 18. A E C D | 19. A E C D | 20. A B C D |

.../

- | | | | |
|-------------|-------------|-------------|-------------|
| 21. A B C D | 22. A B C D | 23. A B C D | 24. A B C D |
| 25. A B C D | 26. A B C D | 27. A B C D | 28. A B C D |
| 29. A B C D | 30. A B C D | | |

PART B

- | | | | |
|-------------|-------------|-------------|-------------|
| 31. A B C D | 32. A B C D | 33. A B C D | 34. A B C D |
| 35. A B C D | 36. A B C D | 37. A B C D | 38. A B C D |
| 39. A B C D | 40. A B C D | 41. A B C D | 42. A B C D |
| 43. A B C D | 44. A B C D | 45. A B C D | 46. A B C D |
| 47. A B C D | 48. A B C D | 49. A B C D | 50. A B C D |
| 51. A B C D | 52. A B C D | 53. A B C D | 54. A B C D |
| 55. A B C D | 56. A B C D | 57. A B C D | 58. A B C D |
| 59. A B C D | 60. A B C D | | |

Appendix 13

Answer key and marking guidelines for external written test

Answer key

Please put '✓' if correct and 'X' if wrong against each item.
1 mark each (100 marks).

;

Section 2: A) 1 mark each (15 marks).

1. C	2. D	3. A	4. D	5. B
6. C	7. B	8. D	9. C	10. A
11. B	12. C	13. D	14. B	15. C

D) 1 mark each (25 marks).

16. C	17. D	18. B	19. D	20. A
21. B	22. B	23. C	24. C	25. A
26. D	27. D	28. A	29. D	30. C
31. C	32. D	33. C	34. A	35. D
36. B	37. D	38. D	39. C	40. A

Section 3: A) 1 mark each (30 marks).

1. D	2. D	3. A	4. A	5. D
6. C	7. B	8. C	9. B	10. D
11. A	12. D	13. D	14. A	15. C
16. A	17. B	18. D	19. D	20. D
21. A	22. D	23. D	24. D	25. B
26. A	27. C	28. B	29. D	30. D

B) 1 mark each (30 marks).

31. C	32. C	33. D	34. D	35. A
36. D	37. D	38. A	39. D	40. C
41. A	42. C	43. D	44. A	45. D
46. A	47. D	48. C	49. D	50. A
51. C	52. D	53. D	54. C	55. A
56. D	57. D	58. A	59. D	60. C

.../

Appendix 14

Answer sheet for external listening test

LISTENING COMPREHENSION

TEST 16

Student's Name; _____

I.D.No. _____

Faculty; _____

Department; _____

Programme (Degree/Diploma); _____

Section; _____

General Direction;

There are four parts in this question paper. Each part of the test will be heard twice. There will be pauses to allow you to look through the questions before each part, and other pauses to let you think about your answers.

Further instructions will be given on the recording. Your answers must be written in this booklet in the spaces provided.

-2-

FIRST PART

You will hear a radio discussion about adopting children. For questions 1-8 tick (✓) one of the boxes to show whether each statement is true or false.

1. She says that adopted children cause problems for grandparents.
2. She believes that most problems are the same for biological and adopted children.
3. She says it is difficult to adopt in Britain if you are over forty.
4. She mentions another country where children over 14 cannot be adopted.
5. The speaker says that she was adopted herself.
6. She thinks that children need a mother more than a father.
7. She describes the case of a child who was adopted after running away from home.
8. She believes that very young children should be told they are adopted.

True	False

-2-

SECOND PART

You will hear a local radio news bulletin. For questions 9-12 tick (✓) one of the boxes A, B, C, or D to show the correct answer.

9. Why do the police want the speed limit changed?

- A. Road Safety experts recommend it.
- B. The present limit is too high.
- C. It would suit modern motorways.
- D. Accident injuries would be less severe.

A	
B	
C	
D	

10. What were the students protesting about?

- A. The use of force by local police
- B. The amount of money students receive
- C. The closure of the Brimcuth-Christchurch road.
- D. The low Standards of their College

A	
B	
C	
D	

11. Why is work being done on the theatre?

- A. To win back public support
- B. To prevent it from falling down
- C. To increase the number of seats
- D. To turn it into Council offices

A	
B	
C	
D	

12. What does the doctor call for?

- A. Smokers to buy 70 per cent fewer cigarettes
- B. a national no-smoking day
- C. no more tax-free cigarettes
- D. no more imported cigarettes

A.	
B	
C	
D.	

- 4 -

THIRD PART

You will hear a girl being interviewed for a job at a swimming bath. For questions 13 to 19 fill in the missing information on the application form below. Some of it has been filled in for you.

Application for the post of Swimming-Bath (13) _____

(14) Part-time Full-time Tick (✓) the appropriate box

SURNAME (15)

FIRST NAMES (16)

AGE ..21.....

LIFE SAVING QUALIFICATION(S) *Exchange medal* DATE OBTAINED (17)

EXPERIENCE (18)

(19) TIMES AVAILABLE FOR WORK Tick (✓) the appropriate boxes

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
orning							
afternoon							
evening							

-5-

FOURTH PART

You will hear a man talking about his job as a guitar teacher. For questions 20-23 tick (✓) one of the boxes A, B, C or D to show the correct answer.

20. Chris teaches children the guitar

- A in one school.
- B privately
- C in his free time.
- D in more than one school.

A	
B	
C	
D	

21. Chris thinks the best pupils are

- A adults.
- B young children.
- C adults around 18.
- D Children of 10 upwards.

A	
B	
C	
D	

22. Some children learn to play the guitar because their parents

- A play too.
- B think it's easy.
- C force them.
- D think it's useful.

A	
B	
C	
D	

23. Chris finds guitar teaching

- A often tiring.
- B occasionally boring.
- C offers little satisfaction.
- D takes too much time.

A	
B	
C	
D	

Appendix 15

Answer key and marking guidelines for external listening test

Test 10*Answer key and marking guidelines

Please put ' ✓ ' if correct and ' X ' if wrong against each item (30 marks).

First part: 1 mark each (8 marks).

- | | | | |
|----------|----------|----------|----------|
| 1. False | 3. True | 5. False | 7. False |
| 2. True | 4. False | 6. False | 8. True |

Second Part: 1 mark each (4 marks)

- | | | | |
|------|-------|-------|-------|
| 9. C | 10. B | 11. A | 12. C |
|------|-------|-------|-------|

Third part: 14 marks total; no penalty for spelling and capitalization errors.

- | | |
|---|---|
| 13. Attendant (1mark) | 16. Helen Mary (2 marks) |
| 14. Part time ✓ (1mark) | 17. 1983 (1 mark) |
| 15. Smith (1 mark) | 18. Beach life guard (last summer?) (1mark) |
| 19. 7 marks, ie, 1 mark for each day (M-S). | |

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun
Morning	✓	✓	✓	✓	✓	✓	
afternoon	✓	////	✓	✓	✓	✓	
evening	✓	////	✓	✓		✓	

Note: If 2 of the boxes for each day is correct award it 1 mark, and if only 1 of the 3 boxes is correct, mark it wrong

Fourth part: 1 mark each (4 marks)

- | | | | |
|-------|-------|-------|-------|
| 20. D | 21. D | 22. C | 23. E |
|-------|-------|-------|-------|

Appendix 16

Student Questionnaire for the Written Test (SQ1WT)
SQIWP

1. Background Information

Direction: Answer items 1.1-1.11 below by filling in the blank spaces with the appropriate responses.

- 1.1. Name of student: _____
- 1.2. Age: _____
- 1.3. Sex: _____
- 1.4. Department: _____
- 1.5. Name(s) of secondary school(s) attended:
 - 1.5.1. _____ Town _____
 - 1.5.2. _____ Town _____
 - 1.5.3. _____ Town _____
- 1.6. ESLCE English grade: _____
- 1.7. ESLCE GPA: _____
- 1.8. Freshman English 101A exam grade: _____
- 1.9. Freshman English 101B exam grade: _____
- 1.10. Semester GPA: _____
- 1.11. Cumulative GPA: _____

2. Regarding the quality of this test as a whole

Direction: Please indicate your opinion about this English Language test by circling the number representing your view (items 2.1. - 2.11).

strongly disagree	disagree	neutral	agree	strongly agree
1	2	3	4	5

Example: My English teacher is
a hard worker

1 2 3 4 5

- 2.1. Writing this test is enjoyable. 1 2 3 4 5
- 2.2. This test assesses or measures my language ability. 1 2 3 4 5
- 2/3. This test has caused me fear, anxiety or disappointment.. 1 2 3 4 5

.../

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	2	3	4	5

- 2.4. This test is (positively) different from other English Language examinations like the ESLCE and/or Freshman examinations I have taken. 1 2 3 4 5
- 2.5. The instructions or directions in this test are clear. 1 2 3 4 5
- 2.6. The format or the lay out of this test is good and convenient. 1 2 3 4 5
- 2.7. The time allotted for this test is fair or just about right. 1 2 3 4 5
- 2.8. The length of this test is reasonable or just about right. 1 2 3 4 5
- 2.9. The passages and the sources of information in this test are interesting. 1 2 3 4 5
- 2.10. The passages and the sources of information in this test are simple enough for Freshman students. 1 2 3 4 5
- 2.11. The grade I will get from this test will be fair in that it should closely reflect my language ability. 1 2 3 4 5
- 2.12. If you have any particular comment (anything good or bad about the test), please specify.
-
-
-
-

3. Regarding the quality of each section or component

Direction: Please rate the quality of each section or component as a test item or as a source of information (eg. passages, etc.) for the test. Now circle the number representing your response (items 3.1 - 3.20).

	very bad 1	bad 2	neutral 3	good 4	very good 5
<u>Example:</u> My English teacher	1	2	3	4	5
3.1. The reading passage, pp. 1-2	1	2	3	4	5
3.2. Section A (items 1 - 3)	1	2	3	4	5
3.3. Section B (items 4-7)	1	2	3	4	5
3.4. Section C (items 8-12)	1	2	3	4	5
3.5. Section D (items 13-17)	1	2	3	4	5
3.6. Section E (The library catalogue cards, P.5)	1	2	3	4	5
3.7. Section E (items 18-22)	1	2	3	4	5
3.8. The dictionary page, P.6.	1	2	3	4	5
3.9. Section F (items 23-27)	1	2	3	4	5
3.10. Section G (Your role, P.7)	1	2	3	4	5
3.11. Section G (the course offering, P.8)	1	2	3	4	5
3.12. Section G (Completing registration form, P.9)	1	2	3	4	5
3.13. Section H (items 28-37)	1	2	3	4	5
3.14. Section I (items 38-42)	1	2	3	4	5
3.15. Section J (the paragraph P.12)	1	2	3	4	5
3.16. Section J (items 43-44)	1	2	3	4	5
3.17. Section k (the diagram, P.13)	1	2	3	4	5
3.18. Section k (item 45)	1	2	3	4	5
3.19. Section L (item 46)	1	2	3	4	5
3.20. Section M (items 47-86)	1	2	3	4	5

.../

Appendix 17

Student Questionnaire for the Listening Test (SQ2L)

SQ2L

1. Background Information

Direction: Answer items 1.1 - 1.11 below by filling in the blank spaces with the appropriate responses.

- 1.1. Name of student: _____
- 1.2. Age: _____
- 1.3. Sex: _____
- 1.4. Department: _____
- 1.5. Name(s) of secondary school(s) attended:
 - 1.5.1. _____ Town _____
 - 1.5.2. _____ Town _____
 - 1.5.3. _____ Town _____
- 1.6. ESLCE English grade: _____
- 1.7. ESLCE GPA: _____
- 1.8. Freshman English 101A exam grade: _____
- 1.9. Freshman English 101B exam grade: _____
- 1.10. Semester GPA: _____
- 1.11. Cumulative GPA: _____

2. Regarding the quality of this test as a whole

Direction: Please indicate your opinion about this listening test by circling the number representing your view (items 2.1-2.10).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Example: My English teacher is a hard worker.

	1	2	3	④	5
2.1. Writing this test is enjoyable.	1	2	3	4	5
2.2. This test assesses or measures my ability to listen and understand spoken language.	1	2	3	4	5
2.3. The instructions/directions in this test are clear.	1	2	3	4	5

.../

strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 2.4. The format or the lay out of this test is good and convenient. 1 2 3 4 5
- 2.5. The time allotted for this test is fair or just about right. 1 2 3 4 5
- 2.6. The length of this test is reasonable or just about right. 1 2 3 4 5
- 2.7. The passages in this test are interesting. 1 2 3 4 5
- 2.8. The passages in this test are simple enough for Freshman students. 1 2 3 4 5
- 2.9. This test has caused me fear, anxiety or disappointment. 1 2 3 4 5
- 2.10. The grade I will get from this test will be fair in that it should closely reflect my listening ability. 1 2 3 4 5
- 2.11. If you have any particular comment (anything good or bad about this test), please specify. _____
- _____
- _____
- _____
- _____

3. Regarding the quality of each section or component

Direction: Please rate the quality of each section or component as a test item or as a source of information (eg. passage, etc.) for the test. Now circle the number representing your response. (items 3.1 - 3.4).

very bad	bad	neutral	good	very good
1	2	3	4	5

<u>Example:</u> My English teacher	1	2	3	④	5
3.1. Section A (the partial dictation)	1	2	3	4	5
3.2. The listening passage read out	1	2	3	4	5
3.3. Section B (items 1-2)	1	2	3	4	5
3.4. Section C (items 3-6)	1	2	3	4	5

.../

Appendix 18

Student Questionnaire for the Oral Test (SQ3S)

SQ3s

1. Background Information

Direction: Answer items 1.1 - 1.11 below by filling in the blank spaces with the appropriate responses.

- 1.1. Name of student: _____
- 1.2. Age: _____
- 1.3. Sex: _____
- 1.4. Department: _____
- 1.5. Name(s) of secondary school(s) attended:
 - 1.5.1. _____ Town _____
 - 1.5.2. _____ Town _____
 - 1.5.3. _____ Town _____
- 1.6. ESLCE English grade: _____
- 1.7. ESLCE GPA: _____
- 1.8. Freshman English 101A exam grade: _____
- 1.9. Freshman English 101B exam grade: _____
- 1.10. Semester GPA: _____
- 1.11. Cumulative GPA: _____

2. Regarding the quality of this test as a whole

Direction: Please indicate your opinion about this oral test by circling the number representing your view (items 2.1 - 2.10).

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	2	3	4	5

- | | | | | | |
|---|---|---|---|---|---|
| <u>Example:</u> My English teacher is a hard worker. | 1 | 2 | 3 | ④ | 5 |
| 2.1. Doing this test is enjoyable . | 1 | 2 | 3 | 4 | 5 |
| 2.2. This test assesses or measures my ability to speak in English. | 1 | 2 | 3 | 4 | 5 |
| 2.3. The instructions/directions in this test are clear. | 1 | 2 | 3 | 4 | 5 |

.../

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	2	3	4	5

- 2.4. The arrangements and conditions of this test are good and convenient. 1 2 3 4 5
- 2.5. The time allotted for this test is fair or just about right. 1 2 3 4 5
- 2.6. The length of this test is reasonable or just about right. 1 2 3 4 5
- 2.7. The sources of information (eg. schedules, etc.) in this test are interesting. 1 2 3 4 5
- 2.8. The sources of information (eg. schedules, etc) in this test are simple enough for Freshman students. 1 2 3 4 5
- 2.9. This test has caused me fear, anxiety or disappointment. 1 2 3 4 5
- 2.10. The grade I will get from this test will be fair in that it should closely reflect my speaking ability. 1 2 3 4 5
- 2.11. If you have any particular comment (anything good or bad about the test), please specify.

3. Regarding the quality of each section or component

Direction: Please rate the quality of each section or component as a test item or as a source of information (eg. schedules, etc.) for the test. Now circle the number representing your response (items 3.1 - 3.4).

Very bad 1	bad 2	neatural 3	good 4	Very good 5
------------------	----------	---------------	-----------	-------------------

<u>Example:</u> My English teacher	1	2	3	④	5
3.1. Section A (the dialogue you read out with your examiner)	1	2	3	4	5
3.2. Section B (the double dialogue where you choose and respond to your examiner)	1	2	3	4	5
3.3. Section C (the class schedule)	1	2	2	4	5
3.4. Section C (the asking and answering of questions)	1	2	3	4	5

Appendix 19

Language Teachers Questionnaire for the three tests

LTQWP

A) The Written Test

1. Background Information

Direction: Answer items 1.1 - 1.8 below by filling in the blank spaces with the appropriate responses.

- 1.1. Age _____
- 1.2. Sex _____
- 1.3. Nationality _____
- 1.4. Qualification/highest degree _____
- 1.5. Country(ies) degree earned _____
- 1.6. Area of specialization _____
- 1.7. Number of years teaching Freshman English 101A and 101B _____
- 1.8 Academic rank _____

2. Regarding the quality of the written test as a whole

Direction: Please indicate your opinion about this English Language test by circling the number representing your view (items 2.1 - 2.11).

Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	2	3	4	5

- 2.1. This test assesses or measures language ability. 1 2 3 4 5
- 2.2. This test is (positively) different from other English Language examinations like the ESLCE or Freshman Examinations. 1 2 3 4 5
- 2.3. The instructions/directions in this test are clear. 1 2 3 4 5
- 2.4. The format or the lay out of this test is good and convenient. 1 2 3 4 5

.../

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 2.5. The markability of this test is satisfactory. 1 2 3 4 5
- 2.6. The time allotted for this test is fair or just about right. 1 2 3 4 5
- 2.7. The length of this test is reasonable or just about right. 1 2 3 4 5
- 2.8. The passages and the sources of information in this test are interesting. 1 2 3 4 5
- 2.9. The passages and the sources of information in this test are simple enough for Freshman students. 1 2 3 4 5
- 2.10. The grade the candidate gets from this test will be fair in that it closely reflects his/her language ability. 1 2 3 4 5
- 2.11. Overall, this test is suitable for Freshman students. 1 2 3 4 5
- 2.12. If you have any particular comment (anything good or bad about the test), please specify _____
- _____
- _____
- _____
- _____

3. Regarding the content and quality of each section or component of the written test.

Direction: Please indicate your views by circling the number representing your view (items 3.1 - 3.10).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- | | | | | | |
|---|---|---|---|---|---|
| 3.1. Section A (items 1-3) and section J (item 43) test skimming skill and the ability to distinguish main ideas. | 1 | 2 | 3 | 4 | 5 |
| 3.2. Section B (items 4-7) tests scanning skill. | 1 | 2 | 3 | 4 | 5 |
| 3.3. Section C (items 8-12) tests the recognition of indicators in discourse. | 1 | 2 | 3 | 4 | 5 |
| 3.4. Section D (items 13-17) tests the ability to deduce meanings of unfamiliar words from the context. | 1 | 2 | 3 | 4 | 5 |
| 3.5. Section E (items 18-22) and section F (items 23-27) test reference skills, ie, understanding and using reference sources. | 1 | 2 | 3 | 4 | 5 |
| 3.6. Section G (completing a registration form) tests the ability to understand and use information/ instruction to carry out a task. | 1 | 2 | 3 | 4 | 5 |
| 3.7. Section H (items 28-37) and section L (item 46) test structural accuracy. | 1 | 2 | 3 | 4 | 5 |
| 3.8. Section I (items 38-42) tests the ability to use indicators in discourse. | 1 | 2 | 3 | 4 | 5 |

.../

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 3.9. Section J (item 44) and section k (item 45) test the ability to transcod information, ie, from diagrams/Graphs/ tables to writing and vice versa. 1 2 3 4 5
- 3.10. Section M (items 47-86) tests integrated reading comprehension. 1 2 3 4 5

LTOL

B. The Listening Test

1. Regarding the quality of the listening test as a whole

Direction: Please indicate your opinion about this listening test by circling the number representing your view (items 1.1-1.10).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 1.1. This test assesses or measures the ability to listen and understand spoken language. 1 2 3 4 5
- 1.2. The instructions/directions in this test are clear. 1 2 3 4 5
- 1.3. The format or the lay out of this test is good and convenient. 1 2 3 4 5
- 1.4. The markability of this test is satisfactory. 1 2 3 4 5
- 1.5. The time allotted for this test is fair or just about right. 1 2 3 4 5

.../

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

1.6. The length of this test is reasonable or just about right.

1 2 3 4 5

1.7. The passages in this test are interesting.

1 2 3 4 5

1.8. The passages in this test are simple enough for Freshman students.

1 2 3 4 5

1.9. The grade the candidate gets from this test will be fair in that it closely reflects his/her listening ability.

1 2 3 4 5

1.10. Overall, this listening test is suitable for Freshman students.

1 2 3 4 5

1.11. If you have any comment (anything good or bad about the test), please specify.

2. Regarding the content and quality of each section or component of the listening test.

Direction: Please indicate your views by circling the number representing your view (item 2.1-2.3)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 2.1. Section A (the partial dictation) tests the ability to discriminate sounds and write it down accurately. 1 2 3 4 5
- 2.2. Section B (items 1-2) tests the ability to distinguish the main idea and the ability to listen for specific details. 1 2 3 4 5
- 2.3. Section C (items 3-6) tests the understanding of conceptual meanings. 1 2 3 4 5

LTQS

C. The Oral Test

Regarding the quality of the oral test as a whole

Direction: Please indicate your opinion about this oral test by circling the number representing your view (items 1.1-1.10).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 1.1. This test assesses or measures the ability to speak in English. 1 2 3 4 5
- 1.2. The instructions/directions in this test are clear. 1 2 3 4 5
- 1.3. The arrangements and conditions of this test are good and convenient. 1 2 3 4 5

.../

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- 1.4. The markability of this test is satisfactory 1 2 3 4 5
- 1.5. The time allotted for this test is fair or just about right. 1 2 3 4 5
- 1.6. The length of this test is reasonable or just about right. 1 2 3 4 5
- 1.7. The sources of information (eg. schedules, etc.) in this test are interesting. 1 2 3 4 5
- 1.8. The sources of information (eg. schedules, etc.) in this test are simple enough for Freshman students. 1 2 3 4 5
- 1.9. The grade the candidate gets from this test will be fair in that it closely reflects his/her speaking ability. 1 2 3 4 5
- 1.10. Overall, this oral test is suitable for Freshman students. 1 2 3 4 5
- 1.11. If you have any comment (anything good or bad about the test), Please specify. _____
- _____
- _____
- _____
- _____

2. Regarding the content and quality of each section or component of the oral test.

Direction: Please indicate your views by circling the number representing your view (items 2.1-2.3)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

- | | | | | | |
|---|---|---|---|---|---|
| 2.1. Section A (the dialogue) tests the ability to speak (read aloud) with ease and clarity. | 1 | 2 | 3 | 4 | 5 |
| 2.2. Section B (the double dialogue) tests the ability to respond appropriately with ease and clarity. | 1 | 2 | 3 | 4 | 5 |
| 2.3. Section C (asking and answering questions from sources) tests the ability to ask and answer questions with ease and clarity. | 1 | 2 | 3 | 4 | 5 |

Appendix 20

Item results for external written test (WT2)

<u>Section</u>	<u>Item</u>	<u>Facility value</u>	<u>Discrimination index</u>	<u>Item-total correlation</u>
2A	1	0.794	0.217	0.251
	2	0.765	0.391	0.390
	3	0.818	0.348	0.353
	4	0.876	0.152	0.127
	5	0.782	0.326	0.274
	6	0.800	0.435	0.355
	7	0.765	0.435	0.401
	8	0.688	0.500	0.431
	9	0.571	0.565	0.434
	10	0.388	0.630	0.484
	11	0.829	0.239	0.301
	12	0.359	0.239	0.203
	13	0.582	0.696	0.531
	14	0.676	0.370	0.362
	15	0.612	0.457	0.412
Average	-	<u>0.687</u>	<u>0.400</u>	<u>0.354</u>
2B	16	0.741	0.391	0.361
	17	0.724	0.457	0.443
	18	0.500	0.196	0.203
	19	0.847	0.283	0.297
	20	0.612	0.370	0.319
	21	0.253	0.326	0.325
	22	0.629	0.217	0.204
	23	0.524	0.261	0.256
	24	0.865	0.239	0.278

	25	0.271	0.587	0.521
	26	0.724	0.391	0.318
	27	0.335	0.239	0.216
	28	0.524	0.457	0.324
	29	0.453	0.587	0.517
	30	0.388	-0.043	0.015
	31	0.618	0.391	0.369
	32	0.729	0.457	0.412
	33	0.682	0.500	0.450
	34	0.412	0.065	0.047
	35	0.582	0.587	0.457
	36	0.394	0.087	0.065
	37	0.741	0.413	0.454
	38	0.629	0.522	0.441
	39	0.665	0.457	0.361
	40	0.524	0.565	0.452
Average	-	<u>0.575</u>	<u>0.360</u>	<u>0.324</u>
3A	41	0.788	0.457	0.421
	42	0.900	0.152	0.265
	43	0.859	0.261	0.329
	44	0.729	0.217	0.233
	45	0.800	0.326	0.336
	46	0.741	0.348	0.295
	47	0.835	0.152	0.170
	48	0.618	0.565	0.426
	49	0.594	0.174	0.140
	50	0.412	0.435	0.335
	51	0.571	0.326	0.229
	52	0.418	0.261	0.216

53	0.912	0.065	0.122
54	0.606	0.304	0.237
55	0.341	0.413	0.304
56	0.406	0.196	0.132
57	0.447	0.370	0.315
58	0.641	0.478	0.340
59	0.776	0.196	0.199
60	0.576	0.500	0.385
61	0.888	0.283	0.321
62	0.506	0.217	0.205
63	0.459	0.348	0.222
64	0.335	0.348	0.258
65	0.194	0.065	0.058
66	0.394	0.283	0.228
67	0.547	0.543	0.442
68	0.476	0.522	0.404
69	0.488	0.435	0.369
70	0.247	0.043	0.022
Average	-	<u>0.584</u>	<u>0.309</u>
3B	71	0.365	0.478
	72	0.435	0.565
	73	0.494	0.565
	74	0.300	0.283
	75	0.341	0.565
	76	0.341	0.304
	77	0.365	0.587
	78	0.241	0.261
	79	0.353	0.239
	80	0.276	0.348

81	0.112	0.130	0.147
82	0.447	0.565	0.475
83	0.353	0.565	0.483
84	0.341	0.478	0.413
85	0.241	0.370	0.397
86	0.271	0.500	0.466
87	0.200	-0.043	-0.035
88	0.259	0.348	0.293
89	0.306	0.217	0.238
90	0.224	0.304	0.354
91	0.224	0.109	0.145
92	0.071	0.022	0.061
93	0.282	0.326	0.294
94	0.259	0.413	0.435
95	0.224	0.391	0.386
96	0.147	0.217	0.332
97	0.176	0.304	0.385
98	0.235	0.478	0.464
99	0.212	0.435	0.462
100	0.159	0.196	0.291
Average -	<u>0.275</u>	<u>0.351</u>	<u>0.339</u>
<u>Overall mean</u>	<u>0.504</u>	<u>0.348</u>	

Appendix 21

Item results for external listening test (LT2)

<u>Section</u>	<u>Item</u>	<u>Facility</u> <u>value</u>	<u>Discrimination</u> <u>index</u>	<u>Item-total</u> <u>correlation</u>
Part 1	1	0.179	0.146	0.292
	2	0.735	0.512	0.422
	3	0.649	0.415	0.225
	4	0.543	0.512	0.378
	5	0.371	0.366	0.256
	6	0.391	0.488	0.447
	7	0.536	0.293	0.307
	8	0.609	0.439	0.358
Average	-	<u>0.502</u>	<u>0.396</u>	<u>0.336</u>
Part 2	9	0.119	0.220	0.358
	10	0.291	0.463	0.488
	11	0.523	0.439	0.356
	12	0.185	0.049	0.151
Average	-	<u>0.280</u>	<u>0.293</u>	<u>0.338</u>
Part 3	13	0.020	0.073	0.380
	14	0.642	0.415	0.333
	15	0.205	0.463	0.563
	16	0.609	0.488	0.399
	17	0.318	0.463	0.417
	18	0.079	0.244	0.538
	19	0.126	0.439	0.644
	20	0.252	0.683	0.666
	21	0.252	0.561	0.556
	22	0.212	0.659	0.712
	23	0.205	0.610	0.693

	24	0.298	0.707	0.610
	25	0.179	0.488	0.547
	26	0.589	0.073	0.166
Average	-	<u>0.285</u>	<u>0.455</u>	<u>0.516</u>
Part 4	27	0.649	0.317	0.315
	28	0.722	0.341	0.302
	29	0.344	0.512	0.406
	30	0.424	0.195	0.199
Average	-	<u>0.535</u>	<u>0.341</u>	<u>0.306</u>
<u>Overall mean</u>		<u>0.375</u>	<u>0.402</u>	

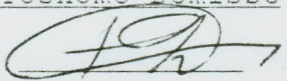
Appendix 22

External listening test (Test 16), cassette (LT2)

DECLARATION

I, the undersigned, declare that this thesis is my work and that all sources of material used for this thesis have been duly acknowledged.

Name: Teshome Demisse

Signature: 

Place: Institute of Language Studies

Addis Ababa University

Date of Submission: June 7th, 1995