

**A Comparative Study of English Language Reading and
Writing Assessment in O Level and Matric System
in Pakistani Private High Schools**



Researcher:

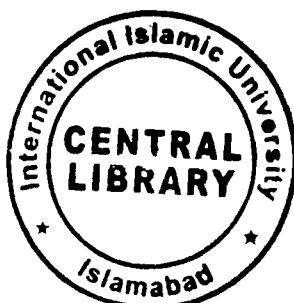
SAIMA HASSAN, née ALI

Registration No: 35-FLL/PHDENG/F09

Supervisor:

DR. FAUZIA JANJUA

**Faculty of Languages, Literature and Humanities,
International Islamic University,
Islamabad, Pakistan.**



TH-16232

Accession No. K
M. M. M.

P^{h^D}
 $U^{2^0} \downarrow$
 SPC

(Accepted by the Viva Voce Committee)

Title of Thesis: "A comparative Study of Reading and Writing Testing System of O Levels and Matric in Pakistani private high schools"

Name of Student: Saima Ali

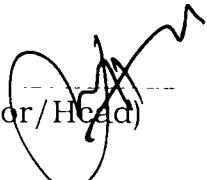
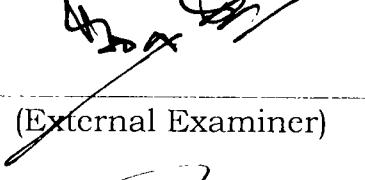
Registration No: 35- FLL/PhD (Eng)/F09

Accepted by the Faculty/ Department of **English (FLL)** INTERNATIONAL ISLAMIC UNIVERSITY, ISLAMABAD, in partial fulfillment of the requirements for the PhD Degrec in **English** with specialization in **Linguistics**.

Viva Voce Committee



(Dean)


(Chairman/Director/Head)
(Supervisor)
(External Examiner)
(External Examiner)
(Internal Examiner)

(September 08, 2014)

ABSTRACT

The present study analyses and compares two high stakes tests and their corresponding testing systems for secondary level in Pakistan i.e. Matric and Ordinary Level (O Level), which are thought to be alternatives of each other. The focus of the study is to determine how the two testing systems are comparable in terms of their English language testing procedures and how the scores from one testing system can be meaningfully compared to the scores from the other.

As claimed by Inter Board Committee of Chairmen (IBCC), the scores from English test from Matric are equivalent to Secondary School Cambridge tests like O Level, General Certificate of Secondary Education (IGCSE) and has an equivalence formula for that. So, in theory, both the assessment systems are comparable and in a way, a claim could be made that students with the same linguistic competence are awarded the same score but in practice this might not be the case as this thesis is trying to demonstrate. Therefore, to see if the scores from O Level and Matric English Language Assessment can actually be compared meaningfully, the researcher using the Common Test Approach compared their previous scores with the scores from the Common Test used in this study. The purpose was to test the English language skills of the participants and to find out whether the test scores in the common test (i.e. IELTS reading and writing modules) correlate with Matric or with O Level, and if yes, with which do they correlate better. On the basis of the current findings, it can be stated with high confidence that the current ratings of O Level grades in this subject are not appropriate, and would suggest that their valuations viz-a-viz the Matric scale should be increased.

Furthermore, linking of Matric English test to Common European Framework of Reference (CEFR) was attempted to see if the Matric could be aligned with an international framework

like CEFR which not only provided evidence about its validity but also helped in comparing it to O Level.

According to the findings of this study, it has been empirically substantiated that the O Level group clearly out performs in the English reading and writing skills of the Matric group on the Common Test. O Level group out performs Matric group by one point on the Common Test (subtest of IELTS) which is significant for the scale of 9 points. In fact, even half a point can restrict a candidate's academic and professional opportunities.

**A Comparative Study of English Language Reading and
Writing Assessment in O Level and Matric System
in Pakistani Private High Schools**

SAIMA HASSAN, née ALI

Registration No: 35-FLL/PHDENG/F09

Submitted in partial fulfillment of the requirements for the
PhD degree in Linguistics
At the faculty of Languages, Literature and Humanities,
International Islamic University,
Islamabad, Pakistan.

ACKNOWLEDGEMENTS

I am grateful to my supervisor who not only guided me through this thesis but whose objectivity and resilience was a beacon for me to accomplish what I have today. I would also like to thank the faculty of English Department at Islamic International University for the care they took in educating me through the degree. Their advice, suggestions and occasional imperatives have contributed immensely to whatever merits my thesis might have.

I profoundly acknowledge the generous support of the Department of Theoretical and Applied Linguistics, University of Cambridge, UK which helped me through my research of this thesis. Thanks to Napoleon Katsos for his support, guidance and consideration as my PhD supervisor without whom this study would not have been possible. Thanks are owed to Chris Cummins for his help and contribution in the statistical analyses of the data.

I would like to thank all the participants who let themselves cajoled into taking part in this study which is reported. Thanks to Ayesha Riaz and other experts for providing their expert opinion when required. I am also thankful to Higher Education of Pakistan for offering me the visiting fellowship in 2012 to the UK which enabled me to broaden my horizon, find out the latest in terms of different language frameworks utilized globally and make it feasible for me to get all the required information about CEFR.

This study is dedicated to my family who has, to their best knowledge and ability, encouraged me to reach the highest level of education possible throughout my life, for which I am greatly indebted to them--most of all, my husband Hassan who has always been my greatest support and strength.

Saima Hassan, née Ali

CONTENTS

List of Figures	xii
List of Tables.....	xiii
Chapter 1	
1.1 Introduction.....	1
1.2 Statement of the Problem	7
1.3 Rationale of the Study	8
1.4 Significance of the Study.....	9
1.5 Objectives.....	10
1.6 Hypotheses.....	10
1.7 Overview of Research Methodology.....	11
Definition of Key Research Terms	13
Chapter 2	
Review of the Literature	
2.1 Introduction	16
2.2 Part 1: Exploring the Need of Testing	17
2.2.1 Assessment and Educational Reform.....	17
2.2.2 Functions of High Stakes Tests	19
2.2.3 Origin of Comparability studies	20
2.2.4 Comparability Studies: Maintaining Standard	21
2.2.5 Comparability studies in Pakistan	24
2.3 Part 2: Exploring the Concept of Testing	25
2.3.1 Language Testing	25
2.3.2 History of Language Testing	25
2.3.3 Theoretical Background to Language Testing and Assessment	27
2.3.4 Assessment/Testing cultures	32
2.3.5 What Does the Language Test Measure?.....	34

2.3.6 Essential Test Qualities	37
2.4 Part 3: Developing Frameworks of Reference for Learning, Teaching, and Assessment	40
2.4.1 Growing Popularity of CEFR	42
2.5 What does the language test score tell?	46
2.5.1 Criteria for Scoring	46
2.5.2 Tests and Language Proficiency Scales	47
2.5.3 Common Reference Levels	48
2.5.4 Can Do Statements.....	52

Chapter 3

Context of the Study

3.1 English Language Testing in Pakistan	55
3.2 Education System of Pakistan.....	56
3.2.1 Qualification System of Pakistan.....	57
3.3 IBCC and the Educational Boards in Pakistan	59
3.3.1 Federal Board of Intermediate and Secondary Education.....	61
3.4 College/ University entrance procedures.....	62
3.5 International English Language System (IELTS).....	63
3.6 Emergence of a National Testing System and its Role in the Education System.....	64
3.7 The Problem: Transparency: Accreditation and equivalence	65
3.8 Language Testing at Secondary Schools in Pakistan	66
3.9 Language Testing at Secondary Level in Pakistan: O Level and Matric	66
3.10 O Level	67
3.10.1 History of O Level	68
3.10.2 Use of O Level in Pakistan.....	68
3.10.3 Key Numbers of O Level	68
3.10.4 Curriculum of O Level	69

3.10.5 O Level Examinations: Approach and Specifications	69
3.10.5.1 Exams and Marking Scheme	70
3.10.5.2 O Level English paper	71
3.10.5.3 Course Objectives and Test Items.....	72
3.10.5.4 Exam Papers	73
3.10.5.5 Paper Setting of O Level	74
3.10.5.6 Marking of O Level.....	74
3.10.5.7 Variability of Examiners/Markers.....	75
3.10.6 O Level: Validation	76
3.10.7 CEFR and O Level.....	78
3.11 Matric	79
3.11.1 History of Matric	80
3.11.2 Use of Matric in Pakistan	80
3.11.3 Key Numbers of Matric	81
3.11.4 Curriculum for English: Matric	81
3.11.5 Matric Examinations: Approach and Specifications	82
3.11.5.1 Exams and Marking Scheme of Matric	84
3.11.5.2 Matric English question papers	85
3.11.5.3 Course Objectives and Test Items.....	86
3.11.5.4 English Exam Papers of Matric	86
3.11.5.5 Paper Setting and Marking of Matric English question papers	88
3.11.5.6 English Paper Setting of Matric	88
3.11.5.7 Marking of Matric English papers	89
3.11.5.8 Variability of Examiners/Markers.....	90
3.11.6 Validation of Matric English Exam	91
3.11.7 CEFR and Matric.....	92
3.11.8 The Problem: Conversion Formula of IBCC	92

Chapter 4

Research Design for the Study

4.1 Research Methodology	94
4.2 Hypotheses.....	95
4.3 Delimitations of the Study	100
4.4 Time Framework for the Study.....	101
4.5 Research Framework.....	101
4.6 Rationale of Using Qualitative and Quantitative Methods.....	103
4.7 Sample of the Study	105
4.7.1 Unmatched group	106
4.7.2 Matched group	107
4.8 Raters.....	108
4.9 Research Tools.....	109
4.10 Decisions on Data Inclusion	112
4.11 Procedure of the Study	112
4.11.1 Phase 1	112
4.11.2 Phase 2	114
4.11.2.1 Procedures and Materials used for Alignment process to CEFR	116
4.11.2.1.1 Familiarisation	116
4.11.2.1.2 Specification	117
4.11.2.1.3 Standardisation.....	118

Chapter 5

Data Analysis

5.1 Phase 1	122
5.1.1 Data Analysis.	122
5.1.2 Linear Regression analysis.....	125

5.1.3 Analysis for matched groups.....	128
5.1.4 CEFR and Matric Scores.....	134
5.1.5 Discussion: The Conversion Scale	136
5.2 Phase 2: Results of Alignment of Matric English Exam to CEFR	137
5.2.1 Familiarisation	137
5.2.2 Standardisation	138
5.2.3 Validation	141
5.2.4 Some Procedural Issues.....	142

Chapter 6

Findings, Conclusions and Recommendations

6.1 Findings	143
6.2 Conclusion	146
6.3 Recommendations: Areas for Follow-up and Further Research	147

Chapter 7

Implications for Matric English Test and Testing Procedures

7.1 A Uniform Change.....	150
7.2 Alignment to CEFR.....	150
7.3 Communicative Approach to Teaching and Testing.....	151
7.4 An Aligned Curriculum	151
7.5 Validation of Examination System	154
7.5.1 Concurrent and Predictive Validity	154
7.5.2 Construct and Content	154
7.5.3 Exam Construction Process.....	156
7.6 Reliability	157
7.7 Examiners Training	158
7.8 Marking Scheme of Matric English Papers.....	158

References	159
Appendices	175
Appendix A	176
Matic English Curriculum (Objectives for Reading & Writing)	
Appendix B	183
O Level English Syllabus (Aims and Objectives)	
Appendix C	193
English Paper: O Level	
Appendix D	204
English Paper: Matic	
Appendix E	211
IBCC Conversion Formula	
Appendix F.....	215
Socio-Economic Questionnaire	
Appendix G	216
CEFR Table A1	
Appendix H.....	217
CEFR Table C4	

LIST OF FIGURES

Chapter 1 Introduction

Figure 1 Relationship between testing, teaching and learning 5

Chapter 2 Review of the Literature

Figure 2 6 Levels Defining Language Proficiency 49

Figure 3 A branching approach 50

Chapter 5 Results and Data Analysis

Figure 4 Scores of Matric and O Level on the common test 124

Figure 5 Average Scores of Matric and O Level on the common test. 125

Figure 6 Linear regression analysis for Matric versus Average score (Unmatched group)
..... 127

Figure 7 Linear regression analysis for O Level versus Average score(Unmatched group)
..... 128

Figure 8 Linear regression analysis for Matric versus Average score (Matched group)
..... 130

Figure 9 Linear regression analysis for O Level versus Average score (Matched group)
..... 131

Figure 10 Distribution of errors on current conversion of IBCC 132

Figure 11 Distribution of errors on proposed conversion 133

Figure 12 O Level grades converted according to the proposed and IBCC formula 134

Figure 13 Mapping of the IELTS scale to the CEFR's six levels 135

Figure 14 Summary of rating Sample Matric English Answer Papers (SSC I) 140

Figure 15 Summary of rating Sample Matric English Answer Papers (SSC II) 141

LIST OF TABLES

Chapter 2	Review of the Literature	
Table 1	Common Reference Levels: global scale.....	51
Table 2	Assessment: positive and negative criteria	53
Chapter 3	Context of the Study	
Table 3	Summary of the desired test validation processes and aspects	78
Table 4	What are the CEFR levels of Cambridge IGCSE English as a Second Language (E2L) by skill?	79
Chapter 4	Research Design for the Study	
Table 5	Unmatched group	107
Table 6	Matched groups	108
Table 7	Material used in the training rating activity	119
Chapter 5	Results and Data Analysis	
Table 8	Scores of Matric and O Level on the common test.	123
Table 9	O Level grades converted according to the proposed and IBCC formula.....	134
Table 10	Relationship of CEFR six levels, IELTS and Matric scores.....	136
Table 11	Conversion of labels representing CEFR levels into a numerical scale	138

FORWARDING SHEET

The thesis entitled A Comparative Study of English Reading and Writing Assessment in O Level and Matric System in Pakistani High Schools submitted by Saima Hassan, nee Ali in partial fulfillment of Ph.D degree in (Discipline) Linguistics has been completed under my guidance and supervision, I am satisfied with the quality of student's research work and allow her to submit this thesis for further process of as per IIU rules & regulations.

Dr. Fauzia Janjua

14th September, 2014

CHAPTER 1

Chapter 1 defines the aims, rationale and research hypotheses. It lays out the research background for the present study to analyse and compare two prevailing testing systems for secondary level in Pakistan. Matric¹ and Ordinary Level (O Level), which are thought to be alternatives of each other in Pakistan, at least for the official purposes. The focus of the study is to determine how the two testing systems are comparable in terms of their English language testing procedures and how the scores from one testing system can be meaningfully compared to the scores from the other.

1.1 Introduction

As is well known amongst education specialists, assessment methods are crucial to educational processes because tests and examinations have a direct impact on the educational processes in various ways (Taylor, 2005). The high-stakes tests² are important according to Madaus (1988) as they bear an important impact on the stakeholders because on the basis of the results from these tests, some important decisions are made directly affecting them. They not only affect the lives of the stakeholders for example students, their parents, teachers but also the practices and policies in the classrooms, schools, education system and society as a

¹ Matric is a term that refers to the final examinations of class 9 and 10 in Pakistan which results in the issuance of Secondary School Certificate (SSC).

² High-stakes tests are those on the basis of which important decisions are made. Offers of future academic and employment opportunities are made on the basis of their results (Cheng, 2005).

whole (Pearson, 1988; Wall, 2005). They may influence the way learners and teachers behave, their opinions about the educational programmes, education system, their approach towards certain educational objectives and activities, most of all learners' and teachers' perception about their own educational worth and capabilities, the academic and job opportunities available to them (Ebel, 1966, cited in Wall, 2005). Thus, all this makes testing³ a powerful tool which serves as a gateway to the land of academic and professional opportunities.

For the aforementioned consequences of important tests, they bear utmost importance for individuals, education systems and societies. Therefore, assessment practices have constantly been undergoing major paradigm shifts in a hope that the successive paradigm will overcome the shortcomings of the previous one and therefore yield standardized testing (Biggs, 1995; Genesee, 1994). So, the quest for better assessment practices and a better future compels the exam providers to change keeping the needs of the test users in view. Research studies reveal that high-stakes testing has been used as a tool to force change the overall education system (Popham, 1987; Frederiksen & Collins, 1989). In the list of countries with well resourced education system where efforts were made to change the educational system through examinations include Hong Kong (see Lam, 1993; Cheng, 1997 cited in Wall, 2005), Israel (see Shohamy, Donitsa-Schmidt & Freeman, 1996 cited in Wall, 2005) and the list of countries with less resourced education system include 14 African countries (see Kellaghan & Greaney, 1992). However, there are also some studies refuting the idea of testing to have any beneficial influence on either teaching or learning (Madaus, 1988).

Eckstein and Noah (1993), while documenting the different functions of high-stakes testing explains the use of testing as a means of control which makes them a powerful tool to

³ The terms tests and examinations are used interchangeably in this thesis.

control educational systems. Wall (2005) believes the potential power of tests can be felt more starkly in the developing countries where because of the fewer vacancies available in the higher education and job market, the stakes of these tests are much higher. Mathews (1985) further supports the idea that the pressure of the selection through tests and examinations is much severe in the developing countries. Foster (1992) also believes that the educational credentials in the less developed world are of much more importance than the developed countries of the world. Therefore, it is for these reasons where high-stakes tests serve as the filter to academia and job opportunities, the teachers adjust their teaching practices according to the test pattern and to suit the test requirements to ensure success in the test where emphasis is more on passing the test than learning the skill.

The intrinsic relationship between teaching, learning and testing is captured by the important phenomenon called *backwash* (Hughes, 1989), now more widely known as *washback* (Weir, 1990; Alderson and Wall, 1993; Alderson, 2004). *Washback* is the impact of a test on learning and teaching (Wall & Alderson, 1993, further explored in Chapter 2). Eisemon (1990) asserts that changes in teaching are only possible if the test designers are ready to co-operate. He further develops the idea by exemplifying his research in Kenya (1988) where teachers were promoting “bad cramming” i.e. class drills with little emphasis on individual study, continuous exposure to possible exam questions and their possible correct answers by the teachers leaving no room for the development of the students’ basic skills. Alderson and Wall (1993) presented 15 washback hypotheses where the impact of testing on teaching and learning was identified. Using Alderson and Wall’s (1993) and Hughes’ (1994 cited in Wall, 2005) research, Bailey (1996) explored the concept of washback and its influence on the educational process and possible ways to promote its positive influence. She also identified the distinction between participants (students, teachers,

material writers, curriculum designers and researchers), processes and products (learning, teaching, new materials and new curricula and research results) to develop her own basic model of washback (see Wall, 2005:53). She elaborated her claim that “tests influence the learner” by using 5 of the 15 Alderson and Wall’s (1993) hypotheses, for instance suggesting that a test will influence learning as what and how the learners learn. A test will influence the rate, sequence, degree and depth of learning. She also provided examples on how the learners prepare for important tests. The examples included practicing test taking strategies, practicing items similar in format to the test, enrolling in test preparation courses and even skipping classes to prepare for the test. Similarly, for the influence of test on teaching, she reinforced the idea by using 6 of the 15 Alderson and Wall’s (1993) hypotheses, for example, a test will influence teaching as what and how the teachers teach. Furthermore, a test will influence the rate, sequence, degree and depth of teaching. A test will, also, influence the attitudes of the test users to the content, method, etc. of teaching and learning. This is further supported by the evidence provided by Kirkland (1971) that sometimes teachers focused more on the examination objectives rather than the course objectives. They, at times, even replaced text books with the examination preparation literature and past papers. The last 4 Alderson and Wall’s (1993) hypotheses include more general consequences of tests, for example they suggest that tests that have important consequences will have washback while the ones that do not have important consequences will have no washback. They further reinforce the idea of washback by stating that tests will have washback on all learners and teachers while the washback effects may be for some learners.

Keeping the above in view, the importance of testing cannot be denied and how they influence the whole educational process. To Hughes (1989) the whole educational process is circular. Ideally, each one of the component i.e. teaching testing and learning should give a

feedback with a motive to improve the whole process. The relationship between testing, teaching and learning can be seen in this simple multidirectional cycle below where each is influenced by the other.

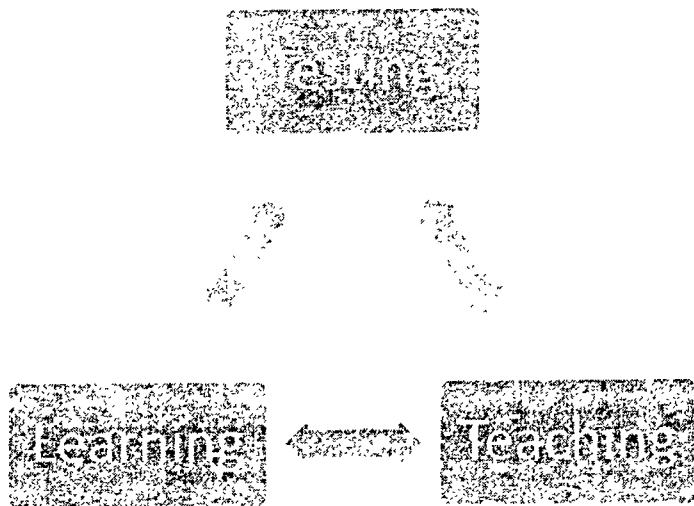


Figure 1: Relationship between testing, teaching and learning

Language assessment, in particular, is a major focus across the world because of increased international mobility and the need for effective communication in academia and professional circles. The aim of language assessment is to infer about an individual's language ability through measurement of his/her latent trait of language ability and the current theories of learning and cognition feed into how this underlying ability can be measured. Any modification or change in these theories causes the need for change in the trends of measuring this latent ability thus, in turn laying grounds for educational reforms. Assessments can be used to bring reform and improve instruction at school level (Linn, 1983, 1992; Noble & Smith, 1994; Popham, 1983, 1987) because assessment and testing play an essential part in educational processes and more importantly they influence the class room pedagogy (Taylor, 2005). So, the relationship of teaching, learning and testing is very pivotal and affects each other gravely.

The scope of this thesis is English language testing so the discussion followed will only cater for English testing. English being an international language has tremendously become popular and its world wide use in the world of academia, jobs, business, tourism etc. is noteworthy. As a consequence, the students are aware of its importance and desire to be competent users of English. Educational reforms, especially in this area are very crucial and required to equip our future generations with the effective tools and strategies to be able to think critically and meet the challenges of the 21st century. In today's democratic societies it is the birth right of every individual to have equal opportunities and having effective linguistic skills ensures a place in the future competitive world. The question arises whether the students are getting fair and equal access to teaching of these skills? and more importantly are they properly assessed and recognized? Teaching and testing are bound to influence each other and therefore have a causal effect on each other as can also be seen in Figure 1. Reforms in one area may possibly automatically improve standards in the other.

The development of frameworks of reference (e.g. Frameworks like the Canadian Language Benchmarks (2000), Common European Framework of Reference (CEFR, henceforth) for Languages developed by the Council of Europe (2001) (see Chapter 2, section 2.4 on Development of Frameworks of Reference for more discussion) is an effort to make these language tests more transparent and interpretable by wider population of test users. With ever increasing mobile populations, qualifications need to be transferable for which it is necessary for the tests to be coherent and transparent. They aim to provide a common scale to describe and interpret levels of language proficiency, facilitating comparison across educational programs, language qualifications, language needs for particular academic and professional purposes, and so on. In 2001, CEFR was produced with the same notion to provide a common scale of making inferences about the testee's latent

ability making the language qualifications more transparent and comparable (North et al., 2010). These frameworks are not only limited to either Europe or Canada but are used widely across the globe (see Chapter 2 section 2.4 for more discussion on this). Many countries have used CEFR to stimulate their language teaching and testing reforms with the goal of improved quality of language education (Martyniuk, 2008).

This research study critically analyses the existing Matric English test by aligning it to an international framework like CEFR which will put forth questions related to the quality of the test thus highlighting its weaknesses and strengths.

1.2 Statement of the Problem

The problem identified in this research is the long felt problem of the stark difference between the English Language Skills of O Level and Matric students even though studying in the same school from the same teachers (referred to excessively in News papers e.g. The News, Dawn, The Times etc.). There seems to be no significant research done regarding this issue, at least in Pakistan to find the reasons of this gap between the two systems i.e. the local Matric stream and the international Cambridge stream. The need to focus on the testing system because of its washback on the education system as a whole is very important and research studies like this one, addressing comparability of examination standards, will prove beneficial in maintaining the standards. The aim of this study is to address important issues of English language assessment/testing in Matric System, more specifically the quality of exam content and the current conversion scale of Inter Board Committee of Chairmen, Pakistan (IBCC, henceforth) in order to contribute towards an improved and more valid and reliable Matric English language testing system at secondary level for the relevant test users, developers and researchers in Pakistan.

Comparing examinations or examination boards and agencies is not new and have been practiced around the globe to maintain standardized education which should be accessible to everyone (see Newton, Baird, Goldstein, Patrick & Tymms, 2007).

The assumption that this thesis makes is that there is a difference between the English reading and writing skills⁴ of O Level and Matric students even if they are from the same school and have the same teachers. The other issue highlighted is the equivalence formula of score conversion from O Level to Matric by IBCC who claims that both these tests are equivalent. For example A* in O Level is equal to 90% of Matric. In order to test this claim, scores from both the tests were compared through Common Test Method and statistically analysed through appropriate statistical procedures (see Chapter 4 for research methodology). Moreover, to further compare the two testing systems, an effort was made to align Matric to CEFR. As O Level is already a recognized international examination, so to see if Matric could also be aligned to an international framework like CEFR, the alignment was carried out.

1.3 Rationale of the Study

This research aims to address important issues of English language testing in Matric System in order to contribute towards an improved and more valid and reliable language testing system for the relevant test users, developers and researchers in Pakistan.

The problem identified in this research is the disparity between English tests of two high-stakes public examinations i.e. Matric and O Level in Pakistan and also compare the English Language Skills of O Level and Matric students to question the conversion scale of IBCC. To emphasize the fact that there is a gulf between the two systems, this research study strives to contribute towards good testing practices in Pakistan which in turn will improve the English teaching methodology. Given that there has not been much research on the

⁴ Since, Matric and O Level English test primarily measures only reading and writing skills of the student.

comparability of examination standards in Pakistan, in spite of many attempts of bringing this issue on the media, there has not been serious and authentic source to claim any of the comparability issues. The researcher felt it very important to come up with some key reasons and rationale to prefer one system over the other i.e. Matric over O Level or vice versa. The question that this thesis raises is that assessment system is the crux of the issue and unless this aspect of education system is controlled properly, the problem will remain static and the disparity between Matric and O Level and the performance of the students from both the systems would even widen with the passage of time. It is time to create a transparent, uniform and coherent system of testing and evaluation of English language in Pakistan in this age of globalization where quality control is the aim of every industry and demand of the masses. This study aims to contribute useful evidence and provide some answers to complex questions about the comparability of exams. It is an attempt to piece together a case about the comparability of the standards of the alternative examinations.

1.4 Significance of Study

Given that comparability of examination standards has been a neglected area in Pakistani education system, results from the following study highlight the English language reading and writing skills of students of O Level and Matric. Matric when aligned to CEFR will make its English test more transparent and coherent and comparable to any English language qualification worldwide. The findings of this research study highlight the English testing practices adapted by Federal Board of Intermediate and Secondary Education (FBISE, henceforth) for administering Matric English test. thus comparing their standards emphasizing the need for more transparent and coherent tests and inferences drawn from them.

Keeping in view the essential traits of a good test (see section 2.3.6), it is believed that Matric and O Level would also be employing the essential qualities in their English Language tests, to what degree, is a matter of investigation. Especially Matric, for which no or very little research is available and therefore this study will attempt to probe into the matter.

This research is also expected to prove beneficial for the policy makers, curriculum designers, textbook authors, teachers, teacher trainers, testers, examination bodies to review the existing Matric syllabus and adapt the content and teaching/assessment methodology and bring it at par to any internationally recognised system. There is a need for all the stakeholders to share the same aims, objectives and assessment criteria.

1.5 Objectives

The aim of the study was threefold:

1. to evaluate and compare the quality of English Language assessment methods and procedures used in both O Level and Matric emphasizing primarily their content validity and scores reliability.
2. to assess the reliability of Matric and O Level English test scores and to see how meaningfully the scores can be compared from Matric and O Level through Common Test Approach and appropriate statistical procedures.
3. to analyse the validity of Matric English testing procedures content through aligning it to an international framework of proficiency i.e. CEFR.

1.6 Hypotheses

The major assumption developed for this study was that the two main streams at secondary level i.e. Matric and O Level, in Pakistan, are generally not evaluating its pupil in the same way and have different testing procedures. The purpose of the study was to see how

meaningfully are the scores comparable from Matric and O Level. Developing a critical outlook towards the existing Matric's English assessment system administered by FBISE, an attempt was made to explore how far they can be aligned to any international framework.

Thus, the following hypotheses were tested in this study.

1. The conversion scale provided by IBCC (see chapter 3, section 3.11.8.1) of converting scores from O Level to Matric undervalues the performance of O Level students by comparison with Matric students.
2. Students of O Level score better on a Common English Test of reading and writing than Matric students.
3. Matric English testing system if compared to English testing system of O Level is not equivalent.
4. English test scores from O Level are more valid and reliable than Matric.
5. Matric English testing procedures practiced by FBISE are not testing the same level of English reading and writing skills of students as O Level by CIE, Cambridge.
6. Matric English testing procedures practiced by FBISE do not possess the same level of validity as O Level when evaluated through the alignment process of CEFR, an international framework, to which O Level already is aligned.

1.7 Overview of research methodology

A combination of quantitative and qualitative research methodology was employed to answer the above research hypotheses. First, the study compared the Matric and O Level scores of the sample with the help of Common Test Approach using quantitative measures (see chapter 4 for detailed discussion of methodology adopted). Then, the second phase of the study with the qualitative methods substantiated the findings put forth in the first phase

by attempting to align Matric English test and testing procedures to CEFR. Many interviews were conducted with the senior officials of FBISE during the course of two years.

Definition of Key Research Terms

The following key terms are operationally defined for the purpose of the present research.

Alignment, according to Fulcher (2010:319) is a “process of linking the meaning of a score to some external criterion, or definition of a level of language ability required for a specific purpose. Content alignment is the process of matching the content of the test and sometimes the content of the curriculum to a set of content standards that state what should be studied and mastered at a particular level of educational achievement.” As a new and a robust phenomenon emerging on the scene of language testing, *alignment or linking* of a language test to any international framework of reference, is making the inferences drawn from these tests more transparent and coherent. It is a process of set of stages aiming to link a test to a framework of reference or to other tests (Cambridge ESOL, 2011). For the present study, alignment comprises of a set of stages chalked out in the CEFR Manual (Council of Europe, 2009).

Comparability, as defined by Newton, Baird, Goldstein, Patrick and Tymms (2007:9) is “the application of same standard across different examinations”. For Pollitt, Ahmed and Crisp (2007) comparability refers to comparing how difficult or demanding a particular examination is than the other. Bramley (2007) yet gives another aspect of comparability by distinguishing between the content standards and performance standards. In this study, comparability deals with the comparison of English testing procedures of O Level and Matric qualitatively. To see how meaningfully both the English tests be compared since both of them are meant for the same level i.e. Secondary level and are supposed to measure the same aspect of English language i.e. reading and writing. The scores from both the tests are statistically analysed to arrive to a confident conclusion.

Validity is a common and one of the most important notion in educational and applied linguistics which means whether the test is measuring what it is intended to measure (Harrison, 1983). The theories presented for the validity of a test may not be as straight forward as many applied linguists would like them to be. It is considered a highly complex phenomenon as can be seen with different definitions for the same phenomenon but from different perspectives. Over the past few decades the theory has changed considerably especially for Messick (1989) among others who believes that it is such an overarching concept beneath all the test related issues can be summarized. Some linguists like Borsboom and Mellenbergh (2004) tend to revert back to a simple definition proposed by Kelley (1927:14) claiming that a test is valid if it measures what it purports to measure and therefore it is not a “complex, multifaceted and dependant on nomological networks and social consequences”. This study also prefers the simpler definition and adopts it from Kelly (1927). This study emphasizes the content validity of the Matric English test. The course objectives are referred to while making a decision whether the test is measuring what it is intended to measure.

Reliability, yet another important phenomenon of applied linguistics and language testing, considering whether the test is consistent and accurate, and therefore dependable (Harrison, 1983). For Bachman and Palmer (1990:19-21), reliability is “consistency of measurement. A reliable test score will be consistent across different characteristics of the testing situation”. Thus, reliability of scores refers to the scores and the inferences deduced from the test scores are reliable and dependable. The present study only probes into the score reliability.

Conversion, according to dictionary definition means “a change from one measuring or calculating system to another...” (Encarta dictionary, 2013). *Conversion formula/scale* is

used by different educational bodies for equivalence of qualifications e.g. Inter Board Committee of Chairmen (IBCC), Australian Average Marks Scaling which converts or rescales the scores from one test to another to make them comparable to a specific exam and for use in other functions (Coe, 2007). In this thesis, it is used to refer to the formula developed by IBCC to convert scores from one examination system to another in order to find an equivalence scale.

CHAPTER 2

REVIEW OF THE LITERATURE

2.1 Introduction

This chapter explores the phenomenon under study *language testing* by capturing its functions, history and evolution over time. It, then, reviews relevant research studies related to language testing highlighting its importance in the educational reform around the world. Common European Framework of Reference (CEFR) is introduced with its significance in promoting good, transparent and fair testing practice around the world.

This chapter consists of the theoretical underpinnings that provided base to the present research study. It consists of three parts. The first part explores the role of testing, its functions and comparability studies in education; the second part traces the origin of tests, what actually the test measures and the essential qualities of a test; the third part of the chapter introduces CEFR, a framework for standardized good practice of language testing and the Manual of CEFR (2009) one of the main research tools to be used in this study. All its main components are described along with the rationale of its selection for the present study.

2.2 Part 1: Exploring the Need of Testing

The first part explores the role of testing in educational reforms around the globe, its functions in a society and how comparability studies encourage standardized testing practices.

2.2.1 Assessment and Educational Reform

It has been established that a good assessment and testing system needs to evolve to keep itself abreast with the changing needs (Alderson & North, 1991). Cheng (2005) believes that educational paradigm shifts are the result of having an incompatible paradigm with the current educational needs and is always seen to overcome the shortcomings of the previous paradigms. To Cheng (2005), paradigm shifts in language assessments are geared by the current theories of learning and cognition with the aim to prepare the students for their future with an increasing success rate.

Testing literature reveal that high-stakes testing is used as a tool to force change the overall education system (Popham, 1987; Frederiksen & Collins, 1989). So, assessment reform Cheng (2005) believes brings an Educational reform as they both are inseparable and are dependent on each other as the desire to pass the test compels the teachers and students to adopt the learning and teaching strategies ensuring the success. Although, Kellaghan and Greaney (1992) after reviewing research on 14 African countries, carried out by the World Bank, carefully pointed out that only examinations cannot change the education system as it is one of the many components in the education system and to have a tangible change there is a need to review the objectives of the educational planners and curriculum developers. But, they also maintain the importance of examinations in an education system.

Thus, change in testing and assessment practices and more importantly its sustainability is of crucial significance for better learning and teaching practices. This fact is

reinforced by the Analysis and Review of Innovations in Assessment (ARIA) project, the planning and implementation of changing assessment practices is important for the teachers, students, governments etc. (Gardner et al, 2008). There are several reform groups striving for better and effective assessment practices such as Assessment Reform Group (being one of the prominent ones) supported by The Nuffield Foundation, Teaching and Learning Research Programme (TLRP) which has been the largest ESRC (The Economic and Social Research Council) programme.⁵ European Association of Language Testing and Assessment (EALTA), Association of Language Testers in Europe (ALTE), American Educational Research Association (AERA) and International Language Testing Association (ILTA) among others are a few prominent organisations working for good quality language testing (Council of Europe, 2009).

Examination bodies around the world are constantly updating their assessment methodologies to provide an accurate measure of the candidates' abilities (see Patrick, 1996 for changes in GCE/GCSE/O and A Levels public examinations). Studies in assessment reforms provide evidence about the need and significance of reforming assessment procedures for better outcomes on national and international level e.g. Eckes et. al. (2005) where they talk about the progress and problems faced in reforming the public language examinations in Europe. According to Cheng (2005), Hong Kong Examination Authorities in 1993 introduced major changes in their major public examination Hong Kong Certificate of Education Examination (HKCEE). However, amid all this heyday of testing reforms, there has also been resistance to change in Language testing, for instance, the protagonists of conservatism believe, as Skehan (1991) points out, that there are some test formats which contain certain wonderful properties that should be retained no matter what. Carroll (1991) identifies the reasons for resisting change which are academic, institutional or financial.

⁵ Retrieved from <http://www.nuffieldfoundation.org/assessment-reform-group> on December 22, 2012.

Pakistani Matric testing system which the researcher believes also needs reforms but might be resisting the change because of either academic, institutional, financial or might even be because of political reasons.

However, this is not the focus of this research study to find the reasons of resisting the change but the aim is certainly to suggest that if not revolution, evolution is certainly needed for quality enhancement of Matric English assessment and testing system which in turn will improve the educational standards. Although, it will not be easy to convince the people/ authorities involved in bringing a positive change but this research will probably prove to be the foundation stone for the more desired change in the field of English language testing in particular.

Comparing two or more testing systems, especially the ones which are considered equal and an alternative of each other assesses their standards. Therefore, to ensure the quality of same qualifications comparability studies are of crucial importance for maintaining standards of tests.

2.2.2 Functions of High-stakes Tests

The functions of high-stakes testing are explored in a comprehensive study by Eckstein and Noah (1993). They explain that the first and foremost function of high-stakes test is that of exercising a control on the entire education system (also discussed in Chapter 1, section 1.1). They document the history of testing and that how testing has been used as a means of control. The second function of high-stakes testing as Eckstein and Noah (ibid) suggest is to check patronage and corruption. Exemplifying Britain, Eckstein and Noah (ibid) documents that before the 19th century access to the higher education and professions was possible through *who* you knew rather than *what* you knew. So, to eradicate corruption examinations were used.

The third function that Eckstein and Noah (ibid) list is to promote higher levels of learning, competence and knowledge. The fourth function was to do justice in allocating the few vacancies in the higher education to the most able pupil. To exemplify this the beginning of 20th century in Japan is quoted. The fifth function according to Eckstein and Noah (ibid) was to improve the teaching standards at school. The last function listed was to minimize the curriculum differentiation in different regions of the country. Again, exemplifying Britain, they state that having a uniform curriculum and thus a centralized education. The establishment of School Certificate examination in Britain were the first ones to have a common goal which were later replaced by the General Certificate of Education (GCE) O Level and A Level.

2.2.3 Origin of Comparability studies

Comparability, to Tattersall (2007), is used as a main driver of maintenance of standards especially in the UK. It emerged as an issue in the mid 19th century in the UK with the development of several educational boards. Since then, its popularity has grown with a set up of a new assessment system in the 20th century and it still continues to shape the educational policies in the 21st century.

A body of research studies traces school examinations like GCE/GCSE/O and A Levels public examinations evolution over time e.g. Patrick (1996) presents an overview of public examinations (GCSE, A Level or GCE) and how meaningfully the standards over time could be compared. There are many discussions on examination policies and practices (Noah & Eckstein, 1989) in international context. Several research studies have also compared international language proficiency tests (see Bachman, Davidson, Ryan & Choi, 1995). Moreover, there is a body of research comparing different aspects of test e.g. Choi, Kim and Boo (2003) compare paper-based and computer-based language tests while Murphy (1982) compares the performance of male and female gender in GCE examinations.

The UK tradition of comparability of examination standards in UK is captured by Tattersall (2007) while some very significant techniques used for comparability are also taken into consideration (Newton, Baird, Goldstein, Patrick & Tymms, 2007).

2.2.4 Comparability Studies: Maintaining Standards

Comparing standards in hope to maintain standards across tests, subjects, testing systems etc. is central to research studies on testing. Setting appropriate standards for the expected testees is yet another issue. Whether the standards are maintained across the country or between boards, it is difficult but not impossible to achieve. Boards have a large body of related comparability research (see Bardell et al., 1978; Forrest and Shoesmith, 1985 cited in Newton, Baird, Goldstein, Patrick & Tymms, 2007).

Every country usually has different traditions in maintaining the standards of language testing as for instance in North America it has always been a common practice to examine the psychometric properties of test where as for UK such practice has established its roots recently. Alderson and Buck (1993) drew attention to how different EFL/ESL Boards in UK did not adhere to any accepted standards and some were not even aware of the concepts like *validity* and *reliability* (see section 2.3.6 on Essential Test Qualities below for the definitions) especially using the psychometric measures. However, there were individual studies e.g. PhD research studies like the present one at that time (i.e. 1993) reported to have been working on the different exams from different boards exploring their validity constructs and so forth. Since then, a large body of research evidence in the UK has gone into the development of psychometric measures of different exams to prove their validity and reliability and the results are available to the test users making the tests more reliable and valid.

Content analysis of different exams usually uses these psychometric measures and has become a norm of good testing practice. The need for content analysis of English exams in Pakistan cannot be denied in order to promote skill learning rather than rote learning as English being a language needs skills rather than mere memorization of vocabulary and set phrases. Pre-testing and psychometric analysis of items can make a high-stakes test like Matric more reliable and valid and such a practice should be encouraged not for fair testing but for promoting better pedagogical practices within the language classrooms (regarding the phenomenon of washback, dealt with later in this chapter in some detail). Such a test, on the basis of which important decisions about the future lives of testees are dependant, needs to adhere to certain standards. It is the responsibility of the test providers to adhere to the highest standards possible and make sure to maintain the highest professional practices and the evidence from it should be publicly accountable too. Studies which fed into the validity of a test provide integrity and fairness to the testing profession. So, to shun such an important issue of validity would be unfair to the test users and certainly not beneficial for the education system in general.

Although, in Pakistan, there is no comprehensive set of standards that the Boards adhere to as each and every board has its own syllabus, book writers, paper setters and examiners etc. So, there is a need to provide them with certain set of rules and measures by the Ministry of education. The ministry should also monitor a uniform standard applied throughout the country for quality and fair education.

Education ministries around the world make sure that uniform standards are applied for fair results. Conditions may not be idealistic as one would like them to be especially in the developing countries but as long as there is will and urge to improve towards more fair and transparent systems in order to be compatible with the more developed systems, there

will be hope to change for the betterment, consequently making the world fair and just with equal opportunities with.

There had been efforts to improve the Matric assessment system in 2005 when a team of specialists from Edexcel, UK visited FBISE, Islamabad, Pakistan⁶. They evaluated the curriculum, text books, papers and the whole process of paper setting and compared the test papers of Matric to O Level and F.A/F.Sc to A Level. On the suggestions and guidelines provided by the testing specialist from Edexcel, some changes were made in the English paper of Matric e.g. there was an effort made to switch to skill based testing rather than memory based. But as predictable, results the following year saw a major decline in the success rate of the students. The failure rate was much higher than the previous years and in comparison with the other boards of the country. Consequently, the affiliated schools got disappointed with the results and started withdrawing their affiliation with the board which forced the board to revert back to the old memory based testing. The new transformed papers with each passing year after 2005 have completely gone back to the old format following memory based testing. The need is to bring a positive change but throughout the country, a uniform change without putting anyone at disadvantage unnecessarily.

There is a range of mechanisms to monitor the adherence to specific standards set by the education ministries but the implementation of any such mechanism for all the examining bodies simultaneously is a challenging job. To monitor the technical quality of examinations and results reliability, validity and comparability of different boards providing the same level of examinations is very crucial for transparent and fair examinations.. Any consistencies found in these areas could lead to unjust results and thus a poor education system. The mechanisms to monitor the technical quality of examinations and results are adopted to a

⁶ Revealed by the officers of FBISE in one of the interviews in December 2012 (see chapter 4. section 4.5).

certain acceptable degree by different educational boards around the world (e.g. for UK see Newton, Baird, Goldstein, Patrick & Tymms, 2007).

2.2.5 Comparability studies in Pakistan

Findings and insights from past language testing research has a major implication for the development and design of the present study. In fact, in a Pakistani context more of the surveys and research studies are related to language policy in education (Rahman, 1997, 1999, 2002, 2007; Mansoor, 2005) but none considering the language testing. Thus, making the investigation of comparisons of high school achievement tests prevalent in Pakistan e.g. Matric a neglected aspect of the education system. To explore this area might give a solution to the issue of the persistent trends of clinging on to the Grammar Translation Teaching Methodology in this modern era of Communicative Language Teaching. Although, there had been a few PhD and MPhil dissertations evaluating the subject papers of Pakistani Matric Examinations but a comparative study of specifically O Level and English testing systems has never been attempted before. The most recent study on the issue is a PhD dissertation (Naeem, 2011) which compares both the systems broadly i.e. teaching objectives, content, methodology and in the end the testing systems. One of the aims of the present study is to investigate the content validity of Pakistani English tests in Matric by attempting to align it to the CEFR which automatically will give rise to any shortcomings if present and the remedies to repair them could be suggested. The exercise of alignment to CEFR will also highlight how much Matric covers more of the desired assessment goals according to an international framework like CEFR. Furthermore, which aspects of a learner's reading and writing skills are assessed by Matric?

2.3 Part 2: Exploring the Concept of Testing

This section contains the basic research phenomenon i.e. language testing, a brief history of its origin, how it developed over time and its significance. It explores how and why language testing is important and what really does it measure and the essential qualities that every good test should have.

2.3.1 Language Testing

Language Testing methods allow the testers to measure through observing the linguistic behaviours of the test takers and evaluate these behaviours through attaching scores which are used as evidence of their inferences (Baker, 1989). The scores, representing the English language ability, then serve as “...the passport of economic prosperity, social mobility and educational advancement” in Fulcher (2007) terms which very aptly echoes the importance of English in this globalised world. As a consequence to keep in line with this growing popularity and demand, new English language measurement instruments are being devised constantly to be used as part of the screening procedures for academic or for work purposes (e.g. classroom placement, university selection, employment selection, professional certification, immigration etc.). For some, as Fulcher (2010) notes, language tests serve as *gate-keeping*⁷ tools.

2.3.2 History of Language Testing

The Biblical evidence of language testing history dates back to far as *Shibboleth* test of Bible quoted in almost all research papers which try to trace language testing history (Kunnan, 1999; Brown and Hudson 2002 cited in Kluitmann, 2008). *Shibboleth* refers to test in which the pronunciation of this word was a passport to life. It was used to differentiate

⁷ “The use of tests to restrict access to education or employment, or to limit international mobility” (Fulcher, 2010:321).

friends from enemies who were destined to die if pronounced it incorrectly. The text quoted in Kluitmann (2008) as:

“And the Gileadites took the passages of Jordan before the Ephraimites: and it was so, that when those Ephraimites which were escaped said, Let me go over; that the men of Gilead said unto him, Art thou an Ephraimite? If he said, Nay; Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand (Judges 12: 5-6)”.

Speaking metaphorically, it is, more or less, still used for the same purpose especially qualification of the international standardized English proficiency tests like IELTS, TOEFL, ICFE etc. which are a means to get into most of the developed Western World.

On the other hand, the history of examinations as Eckstein and Noah (1993) documents to be as old as thousands of years e.g. examinations were used in Imperial China under the Han Dynasty about 200 BC to select their highest ranking officials and probably said to be the first Civil Service Examinations. For the purpose of fair and reliable exams all the scripts were marked anonymously and the final stages were supervised by the Emperor himself (Cheng, 2005). The purpose of these examinations was to break the monopoly of feudals and aristocrats over government jobs. The consequence of this test was substantial and led to control over the educational system where the desire of prospective Chinese candidates to prepare themselves to pass the examinations automatically influenced the teaching (Spolsky, 1995). Eckstein and Noah (1993) state that such a system influenced the development of testing for similar purposes, for example the *Abitur* (1748) by Frederick the Great in Prussia, and the Baccalaureat (1808) by Napoleon in France. Furthermore, in Japan, to lessen the hold of Samurai in the government service, the Meiji emperors also used the competitive examinations and give an opportunity to new talent.

Kirkland (1971) focusing the trends of standardized tests in the United States argues that by the 1960s there were a number of standardized entrance tests in higher education and

many other test batteries used for occupational and professional purposes. As per research statistics in the 1960s, it was estimated that around 150-250 million tests were taken each year in schools alone. The history of modern language system is relatively younger. After the World War I the growing demand of soldiers to have knowledge of foreign languages grew and the US army developed Army Alpha tests. In 1940s, however it became a part of scientific research. According to Kunنان (1999) the first most noteworthy PhD thesis in the field of language tests were Vilareal's Tests of Aural Comprehension (1947) and Lado's Measurement in English as a Foreign Language (1949). Lado's concept of psychometric testing based on discrete point testing and the belief that the language skills can be tested separately was later refuted by the more communicative approach to language testing which catered for the more "interactive, direct and authentic" (Shohamy, 1997: 142) nature of language. The journey from discrete point testing to the more communicative approach to testing is explored below along with the impact of teaching trends on the testing practices of the respective era.

2.3.3 Theoretical Background to Language Testing and Assessment

Throughout the history of learning and teaching, there have been tests and assessments which have been constantly evolving to improve the educational processes and standards. Due to the widespread belief that testing and teaching are interrelated (Harrison, 1983), in the last few decades, the teaching strategies shifted from a more conservative mode of a teacher-centered employing primarily Grammar Translation Method to a more modern, student centered using Communicative Methods especially in language classrooms, thus giving rise to testing procedures to measure these communicative skills more effectively. Around the globe, English Language teaching is moving towards a more target and task

based approach to curriculum and assessment (Cheng, 2005), thus towards a more communicative approach to language teaching and learning.

The changing educational and linguistic theories have had a significant impact on the language teaching over time which in turn had an influence on the language testing. The major reforms in the fields of language teaching and testing are due to the advancement in Theoretical and Applied Linguistics e.g. the contributions by Psycholinguistics, Sociolinguistics, Second Language Acquisition Research, Discourse Analysts, Educational Theory, Validation Theory.

Language testing specialists have to some extent base their development and use of language tests on some theory of language proficiency (see historical reviews of Spolsky, 1995; Weir, 2003). The testing batteries not only have to cater for assessment focus but also are based on some underlying theory about what language and language abilities are.

Spolsky (1976) characterized the historical phases for test development as Pre-Scientific, Psychometrist-Structuralist which for Hawkey (2004) are traditional and modern. Hawkey (2004) believes these phases were influenced by Grammar Translation, Direct and Cognitive code approaches to language teaching and learning. While Psycholinguistic-Sociolinguistic or post modern for Hawkey (2004) emerged in the wake of Communicative Approach to Language Teaching CALT in late 1970s and 1980s and later gave way to Communicative Approach to Language Testing (CALTe). Morrow (1979) translated these eras as *the Garden of Eden* (Pre-Scientific), *the Vale of tears* (Psychometrist-Structuralist) and *the Promised Land* (Psycholinguistic-Sociolinguistic). *The Promised Land*, Morrow (1979) believes was the beginning of the Communicative era. The early models of Psychometrist-Structuralist for language measurement in 1960s were based on the idea of “skills and components” (Lado, 1960; Carol, 1960). Language *skills* i.e. reading, writing,

listening and speaking were distinguished from the language *knowledge* components such as Vocabulary, Grammar, Phonology, Graphology without any indication how the language knowledge components were related to language skills. There was no whatsoever recognition of the role of context in language use. According to Morrow (1979) the most significant features of Psychometrist-Structuralist era are the development of the twin concepts of Reliability and Validity (see below for its definition) as Weir (2003) cites from the design and usefulness of the test that a considerable efforts were put to ensure the validity and useful of what was being tested and the test fairness of test results. Thus, Reliability and Validity are a constant concern of all tests. For Lado (1961) the basis for reliability is objectivity and test of Psychometrist-Structuralist era were influenced by the content framework proposed by Lado (1961). Therefore, Psychometrist-Structuralist test mainly comprised of discrete items that was perceived advantageous in terms of its reliability as these discrete items test could be objectively constructed. Thus, achieving internal consistency and a higher probability of intra- and inter- marker consistency. Such tests Hawkey (2005) believes were also considered to have a degree of validity as the structuralist viewed language and linguistic competence analytically and there was a convenient and direct relationship between the syllabus and the test items.

In 1970's, the importance of context was emphasized by linguists e.g Halliday (1976) draws attention to discourse and language function, Van Dijk (1977) investigates discourse for text/ context relationship and Hymes (1972) explores the sociocultural factors in a situation. In this psycholinguistic-sociolinguistic testing era, Oller (1979) emphasising the role of pragmatics in testing claimed that language proficiency was indivisible and thus could not be tested separately. His beliefs were based on Unitary Competence Hypothesis and

claimed that the testing techniques of pragmatic testing e.g. dictation and cloze tests were able to measure the candidate's language ability more aptly.

Communicative approach to language testing (CALTe) emerged in the 1980s as a reaction against objective tests comprising of Multiple Choice questions due to the perceived over emphasized notion of reliability which to Psychometrist-Structuralists was equal to objectivity. CALTe aims to measure the communicative ability of a candidate as Fulcher (2000) notes that the emphasis of CALTe, is on real life tasks, face validity⁸, authenticity and performance.

Rejecting the notions of *reliability* and *validity* (see section 2.3.6 on Essential Test Qualities below for definitions) of pre-communicative era, CALTe tried to redefine these core concepts of language testing. Canale and Swain's (1980) seminal work, is considered as a foundation stone for CALTe, which in recent years had a great impact on the language teaching and assessment, emphasized the framework of "language competence". They were the first one to propose a communicative competence model that distinguishes grammatical, sociolinguistic and strategic competence. Discourse competence and sociolinguistic competence were further distinguished by Canale in 1983 which presented a communicative model for present day syllabus designers. The "communicative competence" offered a more comprehensive model of the knowledge required for language use than the earlier frameworks of skills and components. Recognizing language use as a dynamic process "language competence" incorporates the knowledge of grammatical rules and how language is used to achieve a particular goal. Candlin (1986: 40) describes "communicative competence" as:

⁸ Face value of the test. A test is said to have face validity if it looks like it is going to test what it is suppose to measure (Hugh, 1989)

“.....a coming together of organized knowledge structures with a set of procedures for adapting this knowledge to solve new problems of communication that do not have ready-made and tailored solutions”.

The other seminal works in language testing worth mentioning are Morrow's (1979), Bachman's (1990). They built more effective tests to measure the test takers language proficiency in context rather than their knowledge about the language as was in the earlier models of skills and components framework in 1960s.

Morrow (1979) emphasizing the qualitative approach to language testing insists that language tests should be able to assess quality of the candidate's linguistic performance rather than assessing quantitatively on how many items were answered correctly. The point that Morrow (1979) makes is that objective tests might not be able to elicit the language competence of the candidate e.g such a test might not be able to elicit any language at all thus the results will be not the true representative of the learners competence. Furthermore, in an objective test the candidate might have to base his/her answers on the examiner's language while in a subjective one the candidate might be able to use his own linguistic repertoire.

Bachman (1990) built more effective tests to measure the test takers language proficiency in context rather than knowledge about the language as was in the earlier models of skills and components framework in 1960s. The concept of communicative competence was further taken up by Bachman (1990) making him amongst the first ones to try and build the elements of communicative competence into a model of Communicative Language Ability (CLA) which was revised in Bachman and Palmer (1996). The model incorporated:

- a) language knowledge i.e. this includes organizational knowledge consisting of grammatical and textual knowledge and pragmatic knowledge that consists of functional and sociolinguistic knowledge.

- b) strategic components i.e. the ability to use components of language competence for a contextualised communication.
- c) psychological mechanisms i.e. neurological and psychological processes involved in the actual execution of language as a physical phenomenon (e.g. processing of perceptual input).

The other aspects which according to Bachman (1996) have an impact on CLA are knowledge structures (1990), topical knowledge (1996), personal characteristics (1996), context of situation (1990), characteristic of languages use or test task and setting (1996) and affective schemata (1996).

In different types of twenty first century language tests whether direct or indirect; discrete-point or integrative; criterion-referenced or norm-referenced; objective or subjective used for whatever purpose e.g. proficiency, achievement, diagnostic, placement (see Hughes, 1989 for detailed discussion on test types) the most common desired aim is to measure the communicative ability of a candidate.

2.3.4 Assessment/Testing cultures

All in all, the trends might have shifted to a more CALTe approach but still the testing and assessment cultures differ in their approaches worldwide. The testing and assessment practices are developed on the basis of social and historical context. In USA, the psychometric tradition and objective test formats dominated as much of foreign language testing in the early twentieth century which had an impact on the practices of language assessment and testing in many parts of the world (Spolsky, 1995). On the other hand, language testing in the UK has always had a strong focus on validation issues and a close relationship between syllabus and curriculum and with pedagogical concerns worldwide

(Weir & Milanovic, 2003). In this tradition the focus is more on task-based testing rather than discrete-point testing.

A body of research can be found on the trends in the assessment practices of specific areas of language e.g. Alderson (2000) on reading; Buck (2001) on listening; Douglas (2000) on language for specific purposes; Luoma (2004) on speaking; Purpura (2004) on grammar; Read (2000) on vocabulary; and Weigle (2002) on writing.

The Cambridge Language Assessment Series (CLAS) also evaluates language testing as a measure of test taker's language competence. Overall, as Kostopoulou (2008) puts it, the current agenda of language assessment is to strengthen the bonding of learning, teaching and assessment keeping in view that all the assessment practices and procedures are ethical and democratic which promote humanistic pedagogies and overall growth and learning of individuals shifting the society to more learner-centered emphasizing the need for lifelong learning.

It is, however, believed now that assessment and testing have a direct impact of on the learning and teaching methodologies. Thus, considering assessment to be integral to student learning, the educationists rely on the feedback from the success or shortcomings of a test and that of a teaching methodology for better standards in teaching and testing. Review of OFSTED (1998, Section 5.6) suggests that the core purpose of testing and evaluation is to improve standards in general and not merely to measure them in particular.

The prominence given to the value of assessment can be recognized from the following quote from an official statement by National Curriculum Task Group on Assessment (TGTA) of UK:

“Promoting children’s learning is a principle aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set and pupil’s progress charted and expressed. It can yield a basis for planning the next steps in children’s needs....it should be an integral part of the

educational process, continually providing both “feedback” and “feedforward”. It therefore needs to be incorporated systematically into teaching strategies and practices at all levels.” (National Curriculum Task Group on Assessment (TGTA): A Report, 1988, cited in Assessment Reform Group, 2000).

2.3.5 What Does the Language Test Measure?

The purpose of assessing language is to measure a latent trait i.e. language ability in order to make inferences about an individual’s language ability. The language testers are able to do this by measuring testee’s behaviour and attach scores to them which are used as an evidence for their inferences. This process is also known as the process of score interpretation (Bachman, 2004). In order to achieve all this successfully, the purpose of the test should be clearly defined and understood. Thus, “fitness for purpose” should be kept in mind while designing the test to make the test useful and similarly the inferences drawn from them more effective.

For the content validity measure, which is, *to what extent the test is measuring what it is meant to* (Harrison, 1983), test developers try to keep in line the curriculum objectives with the subsequent test items. The curriculum designers, keeping in view a specific language learning approach, design a course including all the skills that are desired to be developed. Testing specialists, on the other hand, require to measure the students’ skills and identifying his/her level of competence and for that they need to analyse whether the test items are in line with the course objectives. It is of utmost importance for a reliable and valid measurement to ensure the alignment of the course and assessment objectives. Discrepancy between the course objectives and test content would yield imperfect measurement putting the validity of the test content and scores in question. So, for a valid and reliable test, alignment of course objectives and test content is crucial to measure the students’ capabilities. To achieve this, the curriculum and test developers widely use Bloom’s schemata (1956) of classification of the processes of thinking and learning. Particularly, his

Cognitive Domain got very popular (out of his three domains⁹), with not only the university examiners as initially intended for but all the educationists around the globe. For example, it is extensively used by curriculum planners, administrators, researchers, and classroom teachers at all levels of education (Anderson & Sosniak, 1994). The learner, according to Bloom's Taxonomy, should develop his knowledge and intellect (Cognitive Domain); attitude and beliefs (Affective Domain); and the ability to put physical and bodily skills into effect - to act (Psychomotor Domain). Pollitt, Ahmed and Crisp (2007) assert that since the introduction of O Level examination, it is a common practice to specify the testing content according to the Bloom's Taxonomy (1956) particularly in terms of cognitive skills. Pollitt, Ahmed and Crisp (2007) further elaborate that almost every comparability study compares the intended cognitive demands posited on the testees.

Keeping in view the Bloom's Taxonomy, even without going into the depth of Matric English curriculum and papers, it can be assessed that the English syllabus of Matric is perfectly classified according to Bloom's cognitive skills i.e. higher order cognitive skills are also included in an appropriate proportion. But the subsequent testing items are not aligned with the Bloom's Taxonomy in the same way (see Appendix A & D). The lowest three levels of Bloom's multi-tiered hierarchical model include: remembering, understanding and application. The highest three levels contain more complex skills of analysis, creating and evaluation (Forehand, 2008). Bloom's Cognitive Levels can be classified as:

- **Remembering:** Retrieving, recognizing, and recalling relevant knowledge from memory.

⁹The cognitive - knowledge based domain, consisting of six levels
The affective - attitudinal based domain, consisting of five levels, and
The psychomotor - skills based domain, consisting of six levels.

- **Understanding:** Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
- **Applying:** Carrying out or using a procedure through executing, or implementing.
- **Analyzing:** Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.
- **Evaluating:** Making judgments based on criteria and standards through checking and critiquing.
- **Creating:** Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing (Anderson and Krathwohl, 2001:67-68).

Kellaghan and Greaney (1992) while offering a number of recommendations for improving the examination system suggested that it is very important to assess the higher order cognitive skills to ensure they are being taught. They also insisted that the examination should be a reflection of the entire curriculum and not merely a specific part of it.

Since the Matric English test is mostly memory based, so it is not very difficult to classify all such tasks that are based on memory (see chapter 3, section 3.11.5.2 for content analysis of Matric English paper) according to Bloom's Taxonomy. All such question types fall in the very basic low order thinking skill i.e. *remembering/knowledge* of Bloom's Taxonomy. Matric English paper tends to include more low levels of Bloom's paradigm and the higher ones are ignored therefore one would conclude that an imperfect measurement is yielded with such a test.

2.3.6 Essential Test Qualities

Testing specialists suggest that there must be an appropriate balance between the essential test qualities which are the basis of any good test. Bachman and Palmer (1996) formulated six general qualities of language testing and assessment; reliability, validity, authenticity, interactiveness, impact and practicality. However, Cambridge ESOL considers four of these qualities to be the most important and are also known as (VRIP features); reliability, validity, impact and practicality. For Harrison (1983) *Reliability* of the test is the extent to which test is consistent and accurate, and therefore dependable. *Validity*, on the other hand, is the extent to which the test measures what it is intended to measure.

The validation theory is not as straight forward as many applied linguists would like it to be, it is considered a highly complex phenomenon. Over the past few decades the theory has changed considerably especially for Messick (1989) among others who believes that it is such an overarching concept beneath which all the test related issues can be summarized. Although, Borsboom and Mellenbergh (2004) tend to revert back to a simple definition proposed by Kelley (1927:14) claiming that a test is valid if it measures what it purports to measure and therefore it is not a “complex, multifaceted and dependant on nomological networks and social consequences”.

The usefulness of any test thus depends largely on how the test takers perform on the test and to evaluate this, the empirical investigation of the test performance is very important. According to Bachman and Palmer (1996), to investigate quantitatively the test performance, reliability and validity of the measurement employed and more importantly the test scores, are of crucial significance. For Bachman and Palmer (1996: 19-20), the *reliability* of test scores is “the extent to which test scores are consistent and accurate, and therefore dependable” while the *validity* of test scores is the “the extent to which test scores can be

considered a true reflection of underlying ability". The qualities of reliability and validity have always been presented as in conflict over the testing eras. For Morrow (1979: 147) the types of validity except for the face and possibly predictive validity¹⁰, are all "ultimately circular" thus there is no such thing as "absolute validity". Validity exists in terms of specified criteria which are relative and based upon questionable assumptions e.g. if the basic assumptions or the criteria on which the test is based are challenged, the validity claimed automatically is jeopardized. Thus, in the quest of redefining the core notions of reliability and validity, the early communicative language testers were convinced by Davies (1978) to whom there is always a friction between reliability and validity. So, if a test has more validity, the lesser it would have reliability and vice versa. It is perceived that one can only be achieved at the expense of the other. A more recent approach of language testing experts is that both these qualities are important to test and cannot be evaluated separately, therefore, an appropriate balance should be sought and achieved for overall usefulness of a test.

Taylor (2004) draws attention to the difference between other notions such as *impact* of testing includes the impact of a test more broadly on a society at macro level and at micro level i.e. local or educational level which affects the teaching syllabus, course materials and classroom teaching methodology and is referred to as *washback*. For some language testers *washback* is one dimension of overall test impact, referring to the effects on the educational context (Hamp-Lyons, 1997) while for others *washback* and *impact* are two different concepts relating them respectively to "micro" and "macro" effects within the society. Messick (1989, 1996) considers both as parts of "construct validity".

Washback, therefore, is the "impact of a test on teaching" (Wall & Alderson, 1993) which influences the classroom teaching practices. It refers to the relation between the

¹⁰ Predictive validity according to Hughes (1989) is related to the degree to which a test can predict the future performance of the candidate.

teaching, testing and learning. As mentioned above Hughes (1989) believe that the whole educational process while Davies (1968) believes a good test to be an obedient servant of teaching. Furthermore, Milanovic and Saville (1996) also believe in the notion of language testing to be following language teaching. But the educational phenomenon called *washback* tells a different story. It refers to the way a test affects the teaching materials and classroom management. It believes it is not the testing which follows teaching but it is the other way around, it is teaching that follows testing. This intrinsic relationship of teaching and testing is undeniable and to get a clear picture of evolution of testing procedures, a brief overview of language teaching approaches would be useful to justify the changing trends.

Washback makes it possible to see the effect of a test on teaching on a continuum, stretching from *positive* (beneficial *washback*) on one end to *negative* (harmful *washback*) on the other end. Being one of the important areas of Applied Linguistics, Alderson (1986) was one of the first ones to draw the language testing researchers' attention to the *washback* effect which he thought should be the current focus of language testing research. He believes innovation in the language curriculum could be achieved through innovation in language testing. Davies (1985) suggests that tests should follow the curriculum for a positive washback. Wall (2005) believes that *positive washback* can be achieved if the aims and objectives of the test matched those of the curriculum and syllabus. So, positive washback, she argues would motivate teachers to be guided both by the curriculum and the examinations. Similarly, *negative washback*, Wall (2005) believes is the product of discrepancy between the curriculum objectives and those of the test as they will pull in different directions. Testing literature is littered with studies identifying the impact of testing, for example positive washback (Popham, 1987; Frederiksen & Collins, 1989; Biggs, 1995; Wolf, 1997; James, 2000 and Stiggins, 2001 cited in Wall, 2005) and negative washback

(Madaus, 1988; Frederiksen, 1984; Crooks, 1988; Haladyna, Nolen and Haas, 1991; Smith, 1991; Newstead & Findlay, 1997; Shohamy, 1997; Black & William, 1998 and Zeidner, 1998 cited in Wall, 2005).

Bachman (2004) believes that the evidence of these qualities should be provided by the test developers to the test users. Kellaghan and Greaney (1992) also believe that in order to improve the examination system, each of the examination board should have the capacity to conduct research ensuring the presence of these qualities.

Based on these paradigm shifts in teaching, learning and testing spheres different frameworks were developed over time which are discussed briefly.

2.4 Part 3: Development Frameworks of Reference for Learning, Teaching, and Assessment

In the sections below the general aim and scope of CEFR (as described by Council of Europe, 2001; 2009) will be discussed followed by its impact on language assessment and educational reforms not only in Europe but beyond in other parts of the world.

The development of frameworks of reference is a major topic for discussion amongst the Applied Linguistics circles and need of the hour. Setting standards and conforming to these standards is becoming a trend with the educators worldwide. Frameworks like the Canadian Language Benchmarks (2000), the CEFR for Languages developed by the Council of Europe (2001) aim to provide a common scale to describe and interpret levels of language proficiency, facilitating comparison across educational programs, language qualifications, language needs for particular professional purposes, and so on. These frameworks are not only limited to either Europe or Canada but are used widely across the globe.

CEFR, in particular is a manual intended for language learners, teachers, course and syllabus designers, text book writers and assessing bodies. It provides guidelines on how language is learnt and also gives very useful, adoptable and effective teaching methodologies in order to have coherence and transparency in language education.

CEFR is a framework developed by the Council of Europe developed in 1993-1996 (North, Martyniuk & Panther, 2010) which describes the learners' linguistic abilities with its six broad reference levels in a very comprehensive way in terms of Reading, Writing, Listening and Speaking. The "illustrative descriptors" of CEFR describes what a learner should be able to do at each level in order to communicate effectively. These levels also serve as an assessment criteria measuring the learners' progress at each stage of learning. It provides a common language for the elaboration and planning of language syllabuses, curriculums, examinations, textbooks, assessments etc., across Europe, in parts of Asia and Latin America (see Little, 2007 for detail). CEFR has developed a common basis for explicitly describing objectives for language learning, teaching and especially for assessment. It therefore provides transparency of courses, syllabuses and qualifications, thus promoting international co-operation in the field of modern languages.

Focusing on the real needs of the learners in the real world, CEFR helps to overcome the barriers to communication among professionals working in the field of modern languages arising from the different educational systems of the world. It provides the means for educational administrators, course designers, teachers, examining bodies, etc., to reflect on their current practice and co-ordinate.

With the help of objective criteria for describing language proficiency, it works as a useful tool aiding in the mutual recognition of qualifications achieved in different learning contexts, in different parts of the world which will help in international mobility.

CEFR like any other framework has limitations and room for improvement and in no way prescribe how things should be done but at the moment provides a very effective and useful guide book for learners, teachers, assessors, course designers, examining bodies and educational administrators. CEFR gives a practical meaning to test scores being reported through its scales thus making it possible to compare various language examinations more transparently and straightforwardly. It was initially aimed to benefit Europeans but its impact can be seen beyond Europe.

2.4.1 Growing Popularity of CEFR

CEFR has been translated in 36 languages. More than 20 language testing bodies have conducted formal case studies where language tests are related to CEFR (Council of Europe, 2003). CEFR is being used in countries like Australia, China, Japan, Korea, New Zealand, USA etc. Moreover, it has also been adopted by the United Nations for its teacher training and staff in-service training in all United Nations institutions across the globe (Martyniuk, 2008). The reasons of its popularity are probably “good planning or as lucky timing” (Goullier, 2007:31) or as Van Hest (2007 cited in North, 2009:360) considers it as a useful tool to schools, teachers and students which gives them a platform to:

- “discuss and promote language competence and language learning
- structure and plan the language learning process
- discuss progress and achievements in a transparent way.”

CEFR has been readily adopted in Europe which it was quite evidently meant for but is also being enthusiastically accepted outside Europe (see North, 2009). In Canada it has been formally adopted. Macdonald and Vandergrift (2007, cited in North, 2007:361) put forth the reasons of adoption of CEFR in Canada to be its *construct validity, face validity and contextual validity*.

CEFR provides *Construct validity* as the CEFR level descriptors are based on a theory of communicative competence and are empirically validated. *Face validity* ensured because the CEFR descriptors are harmonious with teachers' perceptions and experiences and with language learners. It also presents *Contextual validity* as the branching approach to both levels and categories is able to cater the "different needs and pedagogic cultures of the provinces."

However, Macdonald and Vandergrift (*ibid*) did articulate their reservations about not having enough differentiation at basic levels for Beginner-level language learners. Therefore, ESOL Cambridge and CEFR are working collaboratively to address the issues concerning the CEFR (Milanovic & Weir, 2010).

Different countries have linked their language tests to CEFR e.g. Taiwan has also linked its national exams to CEFR scale. 16 Spanish speaking countries collaborated to form the International Spanish as a Foreign Language Certification System (SICELE) and provide a common certificate based on CEFR. Osaka University and the Japan Foundation, in USA ELP called "Linguafolio" (Bott Van Houten, 2007) have all tried to align their language tests to CEFR.

The testers across the globe desired to align their language tests to CEFR e.g. the private testing organisations and agencies like Eurocentres, made many requests for guidance after the official publication of CEFR in 2001. Responding to such requests, Manuals for linking tests to CEFR were made available (see Council of Europe, 2003; 2004; 2009 cited in Milanovic & Weir, 2010). These manuals propose the following sets of procedures in order to align any language test to CEFR:

- **Familiarisation:** This stage familiarises the participants with the detailed knowledge of CEFR, its levels and the illustrative descriptors. Different familiarisation activities from a

number of activities suggested in the Manual (2009) can be carried out before and during the face to face interaction with the trainer and trainees. Familiarisation as expressed by Council of Europe (2009) is a pre-requisite for effective linking. It is important to have a familiarization session before each stage of the linking process i.e. specification, standardisation. The features that distinguish one level of the CEFR from the other levels are considered at this stage and the minimum requirements for each CEFR levels are outlined. By the end of familiarization session a considerable and desirable familiarity to CEFR and the descriptors should be achieved.

- **Specification:** To provide the internal and external validity of the test in question this is a self audit stage of the alignment process which investigates the coverage of the examination content and task types thus is a content analysis stage. This stage requires different forms to be filled provided in the Manual (2009). This is an awareness-raising activity about the quality of examinations.
- **Standardisation:** This stage makes sure that there is a common understanding between the participants of the CEFR levels with the help of illustrative samples of spoken and written performances available at www.coe.int/lang. For the purpose of standardized marking, there is a training session as suggested by the Manual (2009) then for the standardisation of the examiners *controlled practice* of scoring is carried out before the *individual assessment* of scripts by the participants. The participants read the samples and score them individually using the appropriate scales provided in the Manual (2009). The scores are then shared and discussed as to why a specific level is awarded before finally consulting the right answer allocated to the already standardized illustrative scripts of written performances.

As generally, the end product of training is the benchmarking of local performances so, after a mutual and clear understanding of the CEFR levels, local samples are related to the CEFR levels and scored by the participants accordingly. Standards are set for the local test and local performance samples.

- **Validation:** As a last stage after familiarisation, specification, standardisation and standard setting, validation is the means to assess quality monitoring, to check whether the preceding stages and activities have been successfully carried out. Aspects of validity and procedures to collect validity evidence are also described in the Manual (Council of Europe, 2009).

The proponents of linking a test to CEFR as a standard believe that such an activity is beneficial to improve the testing system over all. As cited in North (2009), CEFR and the Manual has had an impact on the European institutions by helping them to bring their exams to the international standards by designing them systematically, by defining the performance criteria systematically and defining the cut scores explicitly.

Fulcher (2010), on the other hand, contends conforming of a language test to any such standard. He believes that conforming to standards like CEFR stifles the creativity of the teachers. He argues that such a practice may have vested interests of communities, governments with political purposes. To Fulcher (2010) the use of tests and standards for harmonization results in more and more control and destroys the very essence of creativity. To him eclectic approach should be adopted without harnessing a system through alignment to any set standard.

However, a framework so widely used could be worth looking into and made available to the masses of Pakistan. Keeping in view the “underspecification” and “incompleteness” of the CEFR, as noted by Milanovic (2009), makes it an appropriate and

flexible tool to be applied and adapted to the Pakistani context and suit the local purposes and eventually bring Matric to international standards. CEFR does not prescribe the relevant users but is a tool of guidance to improve language educational standards.

2.5 What does the language test score tell?

To Shohamy (2000), language test scores are the evidence of communicative language ability of the testee. Whatever the criteria of scoring a language test, Shohamy (2000) argues, it should be interpretable by other users e.g. for academic or professional screening procedures. Some evidence on how the testee would use his communicative language ability in a real world tasks, if not all, may thrive to include testing content which is based on authentic material. A carefully thought assessment criteria for the test, familiarizing the examiners with the criteria and the differentiating features of the levels is crucial for reliable scores.

2.5.1 Criteria for Scoring. Different ways of scoring an assessment are outlined by Council of Europe (2001: 189-191), out of which few are listed below.

Testees can be scored with specific defined criteria either *on a scale* (e.g. on a scale of numbers or bands) or *on a checklist* where a candidate is evaluated against a list of points or specific criteria which is thought relevant for a particular level or module.

Holistic judgement is when the assessor makes a global judgement assessing different aspect intuitively while *analytic judgement* is when the assessor judges different aspects separately. Testees can also be judged subjectively, called *impression judgement* or a *guided* one set out with specific criteria. *Impression judgement* would be assessing totally subjectively without any consideration of any specific criteria to a specific assessment but solely on the learner's performance. *Guided judgement*, on the other hand, is where the impression judgement of assessor is complimented with specific criteria.

There is, however, a new trend emerging on the scene of assessment called self assessment. *Self assessment* refers to judgement by one's own self about the proficiency while *assessment by others* means judgement by the teachers, examiners etc.

On the basis from the information gathered from the interviews with the members of the educational board (see chapter 4, section 4.4 for details of the interviews conducted for the research), in Matric the scoring methodology used is predominantly holistic judgement. For some of the questions analytic judgement is used but that too is not very detailed.

The assessment criteria comprises of some sort of different scales against which the testee's level of linguistic competence can be judged. An insight into what the language proficiency scales is given below which would be useful for the understanding of CEFR scales.

2.5.2 Tests and Language Proficiency Scales. Tests have an intrinsic relationship with the Measurement Scales and are useful for all the test users to deduce some meaning out of them. Language proficiency Measurement Scales are called by different names such as “*band descriptors, band scales, profile bands, proficiency levels, proficiency scales, proficiency ratings*” (Alderson, 1991:71) or “*guidelines, standards, levels, yardsticks, stages, scales or grades*” (De Jong, 1992:43 cited in North, 2000). All these attempt to provide “*an ascending series of levels of language competence*” (North et.al, 1992:7) or “*a hierarchy of global characterizations of integrated performance*” (ACTFL, 1986 cited in North, 2000:11).

Scales for language proficiency provide a continuum from a non user of a target language to a complete mastery like a native speaker in the target language. These scales have a description for each level to assign grades to the testees accordingly. North (2000) believes that these scales are the advent of the modern times aiming for more transparent educational systems. Many of these scales as Bachman (1990) puts it are “real-life” approach

to assessment giving the learners and assessors the level of attainment in the real world.

While sketching the origin of measurement scales North (2000) found that the first significant rating scale for language proficiency developed in 1950s was US Foreign Service Institute (FSI scale) followed by the ASLPR (Australian Second Language Proficiency Ratings), the ILR (Interagency Language Roundtable) scale for US government employees and the ACTFL (American Council of the Teaching of Foreign Languages) Proficiency Guidelines.

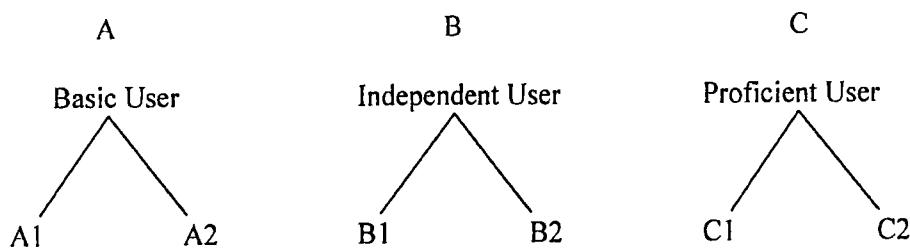
Measurement scales can be of different types e.g. nominal, ordinal, linear, equal interval or ratio scale (see North, 2000 for a detailed discussion). In order to be effective, these measurement scales need to be objective, consistent, should have a clear progression from one level to another, should be comparable with the particulars to be measured and should have a defined zero point to be referred to (Thorndike, 1904; 1912 cited in North, 2000).

Furthermore, there are also different purposes for different scales of language proficiency. See Alderson (1991 cited in Council of Europe, 2001; North, 2000) for distinction between three types of scales such as *user-oriented scale* reporting about “*what the learner can do*” with positively worded descriptors. *Assessor-oriented scale* describing typically “*how well the learner can perform*” with negatively worded descriptors while *constructor-oriented scale* reports on “*what the learner can do*”.

2.5.3 Common Reference Levels. CEFR provides common reference points for language learners and assessors alike. CEFR has six meaningful proficiency reference levels on which Association of Language Testers in Europe (ALTE) examinations are aimed at. These levels were not developed over night but are based on the first reference to a possible set of “Council of Europe levels” based on the proficiency levels proposed by Wilkins

(1978), Trim (1978) and other corresponding levels used by Cambridge Assessments (North, 2000:183).

The six reference levels according to Council of Europe (2001) define the language proficiency as *breakthrough*, *waystage*, *threshold*, *vantage*, *effective operational proficiency*, and *mastery* which are branched from Basic user, independent user and proficient user. It can be illustrated as:



(Breakthrough) (Waystage) (Threshold)(Vantage) (Effective Operational Proficiency)
(Mastery)

Figure 2: 6 Levels Defining Language Proficiency (adapted from Council of Europe, 2001:23)

The branching approach allows the six milestones of language proficiency into further subdivisions giving institutions the flexibility to develop finer distinction between the levels relevant to them to the appropriate degree of delicacy suited to their local needs and system (Council of Europe, 2001; North, 2009). Finnish education system splits A1 into three levels i.e. A1.1, A1.2 and A1.3 as illustrated below in figure 3:

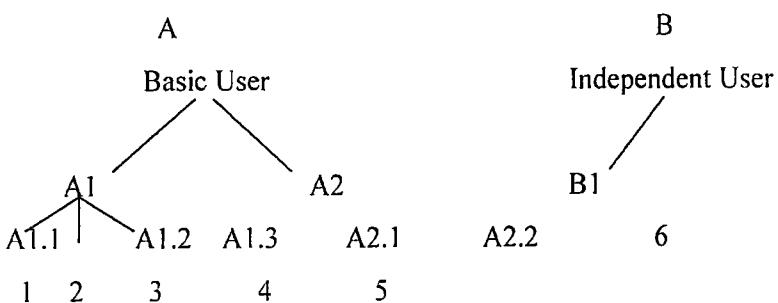


Figure 3: A branching approach adapted from North (2000:358)

Threshold associated with B1 is often recognized as a level required for immigration or citizenship in order to fully integrate in the society (North et.al, 2010).

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without

		strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided.

Table 1: Common Reference Levels: global scale adapted from Chapter 3 of the CEFR (2001:24)

2.5.4 Can Do Statements. To understand where the scales of CEFR come from, an overview of Can Do Statements is very important. Can Do Statements were developed by a long-term research conducted by ALTE. The purpose was to objectively describe exams with in “a framework of key-levels of language performance” (Council of Europe, 2001:244). Can do descriptors provide a medium through which content of the exam or language programme can be related to the rest of the world in terms of what a learner is capable of performing linguistically at a specific level. Previously the descriptors for language assessment were vague e.g. “*Handles routine tasks with an adequate degree of competence*” (North, 2009: 370). It is a common practice for the descriptors especially assessor-oriented to be negatively worded from B2 below. Trim (1978) insists that such descriptors are not very useful as educational objectives and should be avoided. Although, it is a difficult task to formulate positive descriptors especially at lower levels but is desirable to serve as an educational objective. It is preferable to express the same point positively rather than negatively keeping in view the research coming from Applied Psychology. See Table 2 below adapted from Council of Europe (2001:206) for how negatively worded descriptors were developed into positively worded.

Positive	Negative
<ul style="list-style-type: none"> • has a repertoire of basic language and strategies which enables him or her to deal with predictable everyday situations. (Eurocentres Level 3: certificate) • basic repertoire of language and strategies sufficient for most everyday 	<ul style="list-style-type: none"> • has a narrow language repertoire, demanding constant rephrasing and searching for words. (ESU Level 3) • limited language proficiency causes frequent breakdowns and misunderstandings in non-routine situations. (Finnish Level 2)

needs, but generally requiring compromise of the message and searching for words. (Eurocentres Level 3: assessor grid)	<ul style="list-style-type: none"> communication breaks down as language constraints interfere with message. (ESU Level 3)
<ul style="list-style-type: none"> vocabulary centres on areas such as basic objects, places, and most common kinship terms. (ACTFL Novice) 	<ul style="list-style-type: none"> has only a limited vocabulary. (Dutch Level 1) limited range of words and expressions hinders communication of thoughts and ideas. (Gothenburg U)
<ul style="list-style-type: none"> produces and recognises a set of words and short phrases learnt by heart. (Trim 1978 Level 1) 	<ul style="list-style-type: none"> can produce only formulaic utterances lists and enumerations. (ACTFL Novice)
<ul style="list-style-type: none"> can produce brief everyday expressions in order to satisfy simple needs of a concrete type (in the area of salutation, information, etc.). (Elviri; Milan Level 1986) 	<ul style="list-style-type: none"> has only the most basic language repertoire, with little or no evidence of a functional command of the language. (ESU Level 1)

Table 2: Assessment: Positive and Negative criteria

CEFR replaced the negatively worded descriptors into positive ones e.g. “CANNOT answer more than simple, predictable questions” was changed to “CAN answer simple, predictable questions” (Council of Europe, 2001:247).

Can Do statements are useful for learners, teacher, assessor, parents as they give a clear picture what a learner is expected to perform in order to attain a specific level of language proficiency. North (2009) describes the usefulness of *Can Do* descriptors to

teachers to explain the syllabus choice, choose appropriate communicative tasks, learners to see why they are learning certain things. It also helps learners and teachers to set their priorities and assess the progress. It enables the schools to report the learners' progress to parents in a concrete 'salient feature' style. Council of Europe (2001) reports that *Can Do* scales consist of about 400 statements, classified into three general areas: Social and Tourist, Work and Study.

Thus, the descriptors provided by CEFR levels are comprehensive and transparent and in Bachman (1990) words is "real-life" approach to assessment giving the level of testees about their linguistic competence and performance in the real world.

CHAPTER 3

CONTEXT OF THE STUDY

This chapter provides background information about the Pakistani Education system and its secondary school examination systems. The boards offering secondary school certificates, the qualification equivalence organization and its conversion scale, role of English in college and university entrance procedures. The latter part of the chapter overviews the existing secondary school qualification in Pakistan i.e. Matric and O Level. It examines in great detail the international qualification i.e. O Level and its testing system administered by CIE, Cambridge and Matric with its relevant testing system administered by Federal Board of Secondary Education, Pakistan. It explores the testing system; the testing procedures, examination content; papers; marking schemes etc. of both Matric and O Levels.

3.1 English Language Testing in Pakistan

English being the official language of Pakistan is significantly important in all educational institutions in Pakistan. In fact, English has been the medium of instruction in South Asia since 1835 when the Viceroy of British India, Lord Macaulay, laid the foundation stone of a more Westernised education system (Gupta, 2007). It, therefore, has the status of being the language of academia. It is tested at all levels to measure the pupils' linguistic skills to cope in the next academic level. Tests conducted by international examination

bodies are also getting popular such as tests devised by Cambridge ranging from tests designed for secondary schools i.e. O Level, General Certificate of Secondary Education (IGCSE, henceforth) to other English proficiency international tests such as IELTS, TOEFL and other professional examinations like ACCA etc. According to the latest figures available at the official website of the British Council, 420,000 UK exams are delivered in Pakistan by the British Council per annum.¹¹

3.2 Education System of Pakistan

A brief overview of Pakistani schools would be useful here to have a glimpse of the mechanism of education system in Pakistan. According to National Education Policy (1979)¹², Pakistan's education system provides 5 years of free and compulsory education in which English is one of the core subjects. In spite of the efforts, the literacy rate is only 58% according to Pakistan Social and Living Standards Measurement Survey (2011-2012)¹³ because of various reasons, poverty being the number one in the list but since it is not the focus of the thesis, so it is not dealt any further.

Pakistan has private and public schools/ colleges and universities. The public and private schools offer both Matric and Cambridge streams which run parallel. The private schools usually employ the same teachers to teach both the streams. The private and some semi government schools especially the elite schools as termed by Rahman (1997) have adapted a strategy of language teaching in which the medium of instruction and all sorts of interaction is in the target language i.e. English in Pakistani context. This strategy established as an effective and successful form of bilingual education for promoting bilingual education and is known by different names like e.g. *dual language education, two-way bilingual*

¹¹Information retrieved from <http://www.britishcouncil.pk/about/british-council-pakistan> on May 1, 2014

¹²Information retrieved from http://www.finance.gov.pk/survey/chapter_12/10-Education.pdf on January 28, 2013

¹³Information retrieved from http://www.finance.gov.pk/survey/chapter_12/10-Education.pdf on January 28, 2013.

education, two-way immersion, dual immersion, and enriched education (Cloud, Genesse, & Hamayan, 2000; Crawford, 2004; Soltero, 2004). This sort of Dual language programs are not limited to Pakistan but have been implemented in many parts of the world e.g. in the United States for native English speakers and speakers of Spanish, Cantonese, Korean, French, Portuguese, Haitian-Creole, Tagalog, Arabic, and Japanese. Districts have also considered implementing programs in Hmong and Vietnamese (Freeman, Freeman, & Mercuri, 2005). In Pakistan, this form of bilingual education has been successful to different degrees because of the different factors playing their role e.g. the setting of the school, the socioeconomic background, parents' education etc. Bilingual education is usually not found in the public or government schools. Since, to probe into these factors is out of the scope of this study, so it will not be considered any further. Thus, considering the extensive use of English and its role as a medium of instruction in academia, there is bound to be a difference between the English language abilities of the students from both the private and public sector. Now, to assess if there is any difference between the language abilities of the students from the same sector i.e. private, receiving teaching from the same teachers in the same school, the difference is bound to be because of the testing system and because of its washback.

3.2.1 Qualification system of Pakistan. Pakistan has many examination boards, which were working under one central body called Ministry of Education till July 2011 when the ministry was finally devolved and all the boards are working under the control of provincial governments since then, while Federal Board of Intermediate and Secondary Education (FBISE) under the Ministry of Education, Trainings and Standards in Higher Education (official website of Ministry of Education, Trainings and Standards in Higher

Education, 2013)¹⁴. The curriculum used to be developed by the curriculum wing of the Federal Ministry of Education giving clear guidelines for the development of tests and examinations. This curriculum provided to all the provincial boards, however, they also had the liberty to develop their own materials following all the guidelines of the proposed national syllabus/curriculum. The English Curriculum was last developed in 2006¹⁵ by the Ministry of Education then and is still supposed to be implemented by the educational boards across the country. Curriculum contents, therefore, are supposed to be uniform at the national and provincial level.

There has always been skepticism found about the credibility of the examination system of the country especially the high-stakes examinations and the faulty examination procedures are always to be blamed for the deteriorating standards of education. Mansoor (2005) holds the corrupt personnel of government to be responsible for such vicious practices. Because of which rote learning, poor learning and teaching methods are being promoted whilst independent and critical thinking is automatically discouraged. Because of the lack of credibility of the local examinations even the colleges and universities within Pakistan have been using their own entrance exams to supplement or replace results from the current examination system. There is, however, National Testing System (NTS), a government organization, established in 2002 to cater the growing needs of credible results about the candidate's skills and abilities (see section 3.6 on NTS below).

Although, the facade of the English curriculum has a different story to tell (see Appendix A). The English curriculum objectives are as effective and promising as any other international curriculum for the same level. The lapse then comes between the

¹⁴ Information retrieved from :
<http://mopit.gov.pk/gop/index.php?q=aHR0cDovLzE5Mi4xNjguNzAuMTM2L21vcHR0bS9kZWZhdW:0LmFzcHg%3D>

¹⁵ Available at <http://www.share-pdf.com/a79b0aa483ca43669f0109d7ad538c9e/pdf%20f.pdf>

implementation of these curriculum objectives to the actual audience in the classrooms.

Whilst, probing into such questions i.e. finding the gap between the curriculum, teaching materials taught to the pupil in the language classroom and the teaching practices are crucially important for education standards, they are beyond the scope of this study.

3.3 Inter Board Committee of Chairmen (IBCC) and the Educational Boards in Pakistan

The Inter Board Committee of Chairmen (IBCC), an important organization in the education system of Pakistan, is an autonomous organization established in 1972¹⁶, which is responsible for examination reforms, implementation and quality assurance. IBCC, according to its official website¹⁷ (2013), basically ensures the exchange of information among affiliated boards, coordinate activities of the boards of secondary and higher secondary education across the country and aim to standardise academic, evaluation and curricular standards of all the boards. It serves as a channel of all the information and suggestions between the boards and the government. It ensures coordination of educational activities like curriculum, testing and evaluation, examinations, educational planning and administration. Although, IBCC claims to ensure the quality and that the educational boards cannot operate without its approval but the ultimate authority of implementation of curriculum and quality assurance is with the provincial governments and the boards of secondary and higher secondary education at the provincial level and because of that the standard varies from province to province as well as from board to board. This research study only focuses on the FBISE and the sample group is also taken from the schools affiliated with FBISE.

However, there is no evidence present on how IBCC monitors the standards between the boards as one would assume to be done in other parts of the world. As, for example, in

¹⁶Information retrieved from <http://www.ibcc.edu.pk> on May 1, 2014

¹⁷www.fbise.edu.pk

the UK standards across boards are regularly monitored through a number of statistical and judgmental procedures (see Newton, Baird, Goldstein, Patrick & Tymms, 2007).

All the provincial governments e.g. Punjab, Sindh, Balochistan and Khyber Pakhtunkhwa have set their educational boards to conduct matriculation and intermediate examinations. A brief summary of the educational boards of Pakistan is provided below.

In Punjab, a total of eight educational boards are working at present which are Lahore Board, Rawalpindi Board, Multan Board, Dera Ghazi Khan Board, Gujranwala Board, Faisalabad Board, Bahawalpur Board, and Sargodha Board.

Sindh has five educational boards which are Karachi Board, Hyderabad Board, Sukkur Board, Mirpurkhas Board, Larkana Board. There is also Agha Khan University Examination Board (AKU-EB) established in August 2003 works as a National board and provides its services throughout the country. (Official website of AKU-EB, 2013)¹⁸.

Eight educational boards are working in Khyber Pakhtunkhwa namely Peshawar Board, Bannu Board, Swat Board, Malakand Board, Kohat Board, Abbottabad Board, Mardan Board, and Dera Ismail Khan Board.

Balochistan has total three educational Boards that are Quetta Board, Turbat Board, and Zhob Board while in Azad Jammu & Kashmir, there is only one board called Azad Jammu & Kashmir Board. (IBCC, 2012)

Allama Iqbal Open University (AIOU) is another educational board providing distance learning from basic to doctoral programmes since 1974. It also conducts SSC examinations regularly through network regional centres throughout Pakistan.

The above provincial boards affiliate schools all over Pakistan. The criteria set for affiliation of schools are based on certain rules e.g. the facilities available, staff qualifications

¹⁸ Information retrieved from <http://examinationboard.aku.edu/aboutakueb/ebinthecommunity/Pages/ebinthecommunity.aspx>

and educational conditions of the schools. There are also set rules for the academic qualification of the staff to be recruited at school e.g. the principal of a higher secondary school should hold a second class Master's degree (MA/MS), preferably MEd, while a teacher/lecturer should hold a second class Master's degree or an equivalent qualification in the relevant subject (Nordic Recognition Information Centres, 2006).

3.3.1 Federal board of intermediate and secondary education. This study only focuses on the Federal Board of Intermediate and Secondary Education (FBISE) Islamabad because it is one of the largest boards of the country. FBISE is an autonomous body working under the Ministry of Education and Training now after the devolution of the Ministry of Education in 2011. FBISE not only provides services on the national level but also is seen on the international platform as education provider. The institutions affiliated with FBISE are all the schools in Islamabad and several hundred schools across the country including federally administered Northern areas and overseas unlike the other boards which only operate in their respective provinces. FBISE, according to Nordic (2006¹⁹), has about 700 secondary and higher secondary educational affiliated institutions, of which some 50 are based in other Islamic countries especially the Pakistani International schools e.g in Dubai, Abu Dhabi, Saudi Arabia, Egypt, Libya, Iran Qatar, India etc. It has the administrative and financial authority to organise, regulate, develop and control intermediate and secondary education and conduct examinations in the institutions affiliated with it.

FBISE also conducts inspection of the affiliated institutions and according to the FBISE's official website (FBISE, 2013), an affiliation can be withdrawn anytime if the affiliated educational institution does not cooperate and proves to be unable to maintain the standards set by FBISE. The official website of FBISE (2013) also claims that it is

¹⁹ Information retrieved from www.norric.org on 20th, October, 2012

cooperating with the UK board i.e. EDEXCEL in order to have qualifications recognised as UK qualifications (Nordic Recognition Information Centres, 2006).

According to different reports e.g. the 2006 report of Nordic Recognition Information Centres on the system of education in Pakistan highlights that the great differences in quality of secondary and higher secondary education are closely linked to the type of school which corresponds with the socioeconomic class of people e.g. there are semiautonomous schools for the elite as in Rahman's (1997) terms for the affluent sector of the country e.g. the cadet colleges, PAF model schools and other schools like Aitchison school and college (Lahore) Burn Hall (Abbottabad) Gora Gali (Mureec) different Convent schools in the country primarily for girls run by missionary Nuns etc., then there are rural and urban government schools. The report concluded that the rural schools, in particular, lack equipment and trained teachers.

The mission statement of FBISE (2013) claims to be providing efficient, fair and transparent examination system and adhering to international standards.

3.4 College/ University entrance procedures

For admission purposes, evidence of English proficiency is becoming a pre-requisite in many colleges and universities. The following are a couple of examples taken from the official websites of the institutions.

COMSATS Institute of Information Technology, Pakistan states:

“In addition to the academic criteria, proficiency in English Language is also important as the mode of instruction is English. Instead of wasting time and energy in going through the application process, you must self-check whether you meet these criteria.”²⁰

International Islamic University, Islamabad, Pakistan states:

“Arabic and English are the medium of instructions in the University.
..... Ability to read, write and comprehend English is also required. Whereas

²⁰ information retrieved from <http://ciit-atd.edu.pk/admissions/AdmissionProcedure.aspx> on December 18, 2012

in all other faculties (Except Urdu and Persian Departments), English is the medium of instruction. Applicants for admission at IIU are required to pass Language Proficiency Test to qualify for admissions.”²¹

The list can go on as English is the official language of Pakistan, and a pre requisite for admission in the higher education because of its importance on international level. The question arises if the evidence of English language is a pre-requisite for admission in the colleges, so why not make the Matric English exam more valid and the scores more reliable to be used in the screening for admission purposes. This will make the lives of the students easier and the burden on their pockets lighter as the additional test of English proficiency is not free and at times can be as costly as Rs. 19000 for IELTS (May 2013)²². Since, IELTS is very important to this research study, so a brief overview is presented below.

3.5 International English Language Testing System (IELTS)

International English Language Testing System (IELTS) is usually needed for admission in international universities, job opportunities in multinational organizations, migration to English speaking countries, etc., thus claimed by its providers to be an English test for International Opportunity. It holds the same importance for the candidates in Pakistan wishing to study, work or immigrate to English speaking countries. According to IDP (2013)²³, it is the world's most popular high-stakes test. It is delivered in approx 900 centres across the globe taken maximum 4 times a month. It is owned jointly by 3 partners i.e. British Council, IDP; IELTS Australia and Cambridge ESOL examinations (Cambridge ESOL). The official statistics from 2011 reveal that more than 1.7 million candidates sat IELTS around the globe, a world record for any English proficiency test (IDP. 2013)²⁴, which was a 12% growth in number of test in comparison with the number of test takers in

²¹ information retrieved from http://www.jiu.edu.pk/index.php?page_id=1055#mediumhttp://jittatd.edu.pk/admissions/AdmissionProcedure.aspx on December 18, 2012

²² Information retrieved from www.aeo.com.pk on June 7, 2013.

²³ Information retrieved from www.idp.com on June 7, 2013.

²⁴ Information retrieved from www.idp.com on June 7, 2013.

2010 (IELTS, 2013)²⁵. The figures of IELTS test takers reached to 2 million in 2012 (British Council, 2013)²⁶. According to IDP (2013) in 2011 the number of organizations accepting IELTS exceeded than 8000 around the world which included educational institutions, governments, professional associations, employers etc. So, it is potentially a passport to the world of opportunity. Pakistan is considered as one of the top 40 countries where IELTS is taken. Around the year, there are IELTS preparation courses run by a number of government and private educational institutions across Pakistan. The approximate number of candidates from Pakistan was 60000 in 2012 which was shared in an interview with the IELTS providers in Pakistan in July 2013. According to the stats provided by IELTS, in 2012, the mean band score achieved by Pakistani candidates was 5.8 band for Academic and 6.2 band for General Training on a 9-band scale²⁷.

3.6 Emergence of a National Testing System and its Role in the Education System

National Testing System (NTS) is an autonomous government organization, according to its official website (2012), is run by a management committee. It was established in 2002 and has conducted around 3,897 tests since then. These tests are used for screening for college/ university admissions, scholarships, recruitment and promotion purposes. Around 2.93 million candidates sat for test provided by NTS since its inception in 2002. As a pre-requisite of admission in many Pakistani colleges and universities, candidates are asked to provide evidence of their English proficiency and therefore ask them to take one of the NTS tests as can be seen in the example above. Many colleges and universities are thus affiliated with NTS and their candidates take NTS test on regular basis (see section 3.4 above).

²⁵ Information retrieved from www.ielts.org on June 7, 2013.

²⁶ Information retrieved from www.takeielts.britishcouncil.org on June 7, 2013.

²⁷ Information retrieved from www.ielts.org on June 7, 2013.

Many of the tests provided by NTS basically test English proficiency. The products of NTS include: National Aptitude Test, Graduate Assessment Test-General, Graduate Assessment Test-Subject, TOEFL-ITP, A1 English Test (it is an English proficiency test for immigration to English speaking countries which is also authorized by UK Border Agency), Test of English for International Communication (TOEIC).

3.7 The Problem: Transparency: Accreditation and equivalence

To keep up with the modern testing trends, exam boards and other institutions offering high-stakes tests need to demonstrate how valid and reliable their tests and results are and how they are meeting the demands of context. More crucially how do they distinguish between levels of competence in their tests (Weir, 1990). Having a range of available language examinations at regional, national, and international levels, it becomes increasingly important to provide the equivalency tools to the test users. The development of frameworks of reference is therefore, a matter of discussion amongst the applied linguistics circles. Such frameworks include; the Canadian Language Benchmarks (2000), the Common European Framework of Reference for Languages developed by the Council of Europe (2001). Both frameworks aim to provide a common scale to describe and interpret levels of language proficiency, facilitating comparison across educational programmes, language qualifications, language needs for particular professional purposes, and so on. These frameworks are not only limited to either Europe or Canada but are used widely across the globe (see Chapter 2, section 2.4.1). The development of these frameworks is beneficial in a way that it encourages collaboration between language testers and other practitioners such as teachers, admissions staff, employers etc., in addition to making the test results more transparent and comprehensible. These frameworks also allow for comparison of proficiency levels in languages other than English (see Association of Language Testers in Europe cited

in Taylor, 2004). Thus, the importance of these frameworks cannot be denied and the role they play towards more transparent, reliable and valid tests.

3.8 Language Testing at Secondary Level in Pakistan

In Pakistan, the disparity amongst the secondary schools is not one of only private/public sector, English/ Urdu medium of education, but also of different assessment systems. Public sector mostly employs Matric system while the private prefers O Level. Matric, administered by Federal and Provincial Boards under Ministry of Education, and O Level, administered by Cambridge International Examination (CIE) and Edexcel International. These are two secondary level testing systems, used currently to measure the performance of Secondary level students in Pakistan. Matric and O Level are both predominantly (end of year/summative) exam-based and their English language tests are designed to assess many of the same language abilities; reading and writing. These qualifications are usually taken at the age of fourteen-sixteen years. According to the latest figures provided by the FBISE officials in an interview with the researcher, 64327 students took SSC in the year 2012 conducted by only one board i.e. FBISE. CIE, on the other hand, provides examinations and support to around 500 schools in Pakistan (CIE, 2013²⁸). Edexcel is also administering its tests and many schools are affiliated with it. British Council, Pakistan could not provide the information on the number of test takers for O Level because of its secrecy policy. Both the testing systems and their testing procedures are dealt in detail in later sections of this chapter.

3.9 Language Testing at Secondary Level in Pakistan: O Level and Matric

The question arises, when having two different assessment systems in one country, is one preferred over the other? Is there any monitoring for the standardisation of practice and procedures of both the testing systems? Which aspects of a learner's reading and writing

²⁸ Information retrieved from http://www.cie.org.uk/news/announcements/detail?announcement_id=52513 on June 20, 2013.

skills are assessed by each system? And which of the two systems covers more of the desired assessment goals according to international frameworks like CEFR? Moreover, how comparable are the assessments from both the testing systems. It is expected that the results from the following study will be able to highlight the similarities and differences between the English language testing/examinations of O Level and Matric, and consider whether they constitute weak or strong points. They are both high-stakes examinations as the students' future academic and professional opportunities depend on the scores from these exams. Thus, the potential effects of these scores are very significant having a direct impact on the pupils' lives. Evidently, these tests should both be standardized. Since, O Level is an international test and its popularity around the world speaks for its standardised testing practices but the credibility of Matric might be questioned and will be scrutinised in the subsequent sections.

3.10 O Level

Ordinary Level (O Level) is the secondary school examination, widely used in many Commonwealth countries in which Pakistan is also included. It is also used in some other parts of the world like Singapore, Brunei and Mauritius. It is a qualification conferred as part of the General Certificate of Education (GCE). O Level is administered by Cambridge International Examination (CIE) and Edexcel International. O Level, as claimed by CIE²⁹ is especially designed for an international market making the curriculum more sensitive to the needs of different communities and regions which makes the teaching more contextualized and appropriate, therefore more appropriate for students around the world. After O Level the student progresses to Advanced Level (A Level).

²⁹ Information retrieved from <http://www.cie.org.uk/programmes-and-qualifications/cambridge-secondary-2/cambridge-o-level/curriculum/> on June 20, 2013.

3.10.1 History of O Level

In UK, the change of examination suites is a more common trend than in many countries like Pakistan. Research studies feeding into educational research proposes changes to the national and international examinations offered by the boards which are carried forward. In the hope to have a uniform curriculum and thus centralized education in Britain, in the early 21st century the School Certificate examinations were established with a common aim (Eckstein & Noah, 1993). These School Certificate examinations were replaced by the General Certificate of Education (GCE); O Level and A Level. O Level came into existence in 1951 as part of British educational reform in the UK and Northern Ireland. However, in 1988 it was replaced by the General Certificate of Secondary Education (GCSE) and IGCSE which is the international version of GCSE. It can, therefore, be deduced that O Level is not an ultimate best test or it would not have been replaced in the UK but so far in comparison with Matric, it is a more standardized test.

3.10.2 Use of O Level in Pakistan

O Level, in Pakistan, is administered by CIE through British Council which is used currently to measure the performance of Secondary school students. Previously, Edexcel also provided O Level but since 2011 it has been replaced by a new suite of International GCSEs³⁰. So CIE is now the sole provider of O Level in Pakistan. Although, it is slowly being replaced by IGCSE which is another suite of examinations i.e. international GCSE but offers wider range of subject, coursework and syllabus combinations.

3.10.3 Key Numbers of O Level

The number of students taking Matric is obviously much higher than the O Level because of the expenses incurred by O Level which the poor population of Pakistan cannot afford. In Pakistan, a large number of students sit for IGCSE/ O Level conducted by British

³⁰Information retrieved from <http://www.edexcel.com/quals/olevel/Pages/default.aspx>, on January 13, 2013.

Council, Pakistan through extensive networks of centres throughout the country. To have the exact number of candidature, British Council, Pakistan was contacted but no information could be retrieved because of its secrecy policy. However, CIE provides examinations and support to around 500 schools in Pakistan (CIE, 2013³¹).

3.10.4 Curriculum for English: O Level

Considering the needs of students, O Level provides a curriculum up to date with the modern learning, teaching and testing trends. Keeping in line with the CEFR and the modern communicative approach to language learning and teaching, the English curriculum of O Level is based on the communicative approach (see Appendix B for O Level English syllabus). The English curriculum of O Level is especially designed to cater for a wider audience around the globe making use of tasks related to real life situations aiming to develop the communicative competence of the students. These authentic tasks promote meaningful communications which are relevant, challenging, feasible and have identifiable outcomes (Council of Europe, 2001). The English syllabus of O Level aims to make the candidates able to communicate effectively in standard English and enhance their communicative competence, creativity, critical skills and cross cultural awareness. Different resources are used in an O Level class besides the prescribed text books. The students and teachers cannot rely on only one source, as the exam paper content is unseen and if emphasis is laid on any one source can seriously put the success rate of the candidates at risk in the final examinations.

3.10.5 O Level Examinations: Approach and Specifications

The approach adapted in O Level English exam is the communicative approach to Language Testing (CALTe, see chapter 2) as the exam content is not based on the text book

³¹ Information retrieved from http://www.cie.org.uk/news/announcements/detail?announcement_id=5251 on June 20, 2013.

taught in the class but on the syllabus objectives. The question paper includes authentic tasks requiring candidates to use their communicative competence, creativity, critical skills. Being an achievement test, O Level is more subjective in nature than Matric. Subjective assessment according to Council of Europe (2001) as evident from the name is the subjective judgement of the assessor by judging the quality of the performance. Objective assessment, on the other hand, is far removed from the subjective judgement by which the testee gets a mark for answering the right option for a test item e.g. multiple choice questions.

The specifications or as McNamara (2011:31) calls it the “recipe or blue print” for the test construction of O Level are chalked out at the conception of the test which is also reflected in the student booklet giving clear overview of the test content, the length and structure of the test tasks, the type of material used whether authentic or unauthentic, the type of response expected of the candidate, the scoring methods etc.³².

3.10.5.1 Exams and Marking Scheme. As mentioned above (see section 3.4), O Level is exam-based with regular end of year examinations. The English exam for O Level is usually taken in the last year of the Secondary school but private candidates can sit for any of the two examinations held in summer and winter each year. Its English test is designed to assess basically reading and writing of the testees. O Level is an achievement test which is based on a prescribed syllabus. For O Level a wide range of subjects are available and within each subject area there is often a choice of several syllabuses depending on the school and the faculty available. English being one of the core subjects is a compulsory component of the examination. O Levels are generally taken in 7-10 subjects across the curriculum after 11 years of education. The courses usually require 150 hours of teaching as per the CIE official website (2012).

³² See http://www.ciebridgeassessment.org.uk/ca/About_Us/Our_Structure/Exam_Boards/CIE

The exams for O Level are conducted annually in May/June and October/November every year. An O Level student can also re-sit in any of the May/June or October/November session if either because of failing in the previous attempt or if the candidate wants to improve his/her score. In O Level, candidates' performance in individual subjects is assessed on the basis of a six point scale i.e. A*, A, B, C, D, E, where A* is the highest and E is the minimum satisfactory performance and represents a basic level of achievement in a subject and syllabus. Grade U (ungraded), on the other hand, is considered a standard below the lowest acceptable grade i.e. Grade E. In many English speaking countries Grade C satisfies the English language Proficiency requirement of many universities. Each grade in English has a descriptor defining the testee's linguistic abilities in English.

The grades from O Level English exam can be interpreted as an evidence of the candidate's proficiency in English language as they are aligned to the scales of CEFR which clearly states the linguistic ability of the candidate.

The marking key provided to the examiners of O Level relies on analytical marking³³. O Level candidates are told very clearly in their syllabus how they will be assessed (and which areas hold more weightage than the others and so on) with clear grade descriptors. Cambridge CIE has a very comprehensive detail about the specification of O Level, the syllabus aims and objectives, the assessment aims and objectives. Above all, their marking band descriptors for each paper are easily accessible online.³⁴ Previous papers and their detailed marking keys are also available online. Thus, the students have a very fair idea about what and how they will be marked making the exam transparent and reliable.

3.10.5.2 O Level English paper. The meaningful communications promoted through authentic tasks with the identifiable outcomes are tested through similar tasks in O

³³ See chapter 2 for definition

³⁴ At http://www.cambridgeassessment.org.uk/ca/About_Us/Our_Structure/Exam_Blocks/CIE.

Level English paper. The task types are all skill based and requires the candidates to use his linguistic competence to answer the questions. *Writing paper* consists of two sections; *Directed* and *Creative Writing*. In *Directed Writing* a specific response is elicited from the candidate for a specific audience and situation. While, the language is tested in the *Creative Writing* section where the candidate is given a chance to exhibit his/her opinion, creativity, experience and imagination. *Reading*, on the other hand, is tested through *Reading for Ideas* where the candidate scans and summarizes specific information while language is tested through *Reading for Meaning* section which is more demanding and requires the candidate to demonstrate his English language competence through giving more in depth analysis of the text³⁵.

3.10.5.3 Course objectives and test items. As mentioned above, the curriculum of O Level is based on the communicative approach where communicative approach to language teaching and the subsequent testing is employed that are indirectly related to real life situations with the aim to develop the communicative competence of the students. These authentic tasks promote meaningful communications which are relevant, challenging, feasible and have identifiable outcomes. O Level English curriculum equips the pupil with creative and critical thinking, problem solving skills and places great emphasis on the development of skills along with the theoretical knowledge of the subject (CIE, 2013). Since the introduction of O Level in 1951, it is a common practice to specify the content of papers in terms of cognitive skills which are generally derived from the cognitive domain of Bloom's taxonomy of 1956 (Pollitt, Ahmed & Crisp, 2007) discussed earlier in Chapter 2.

³⁵ Information retrieved from CIE website [http://www.gov.mu/portal/sites/mesweb/CIE%20Syllabus%20and%20Support%20Materia%20\(E\)/pdf/1123_y13_sy.pdf](http://www.gov.mu/portal/sites/mesweb/CIE%20Syllabus%20and%20Support%20Materia%20(E)/pdf/1123_y13_sy.pdf) on January 13, 2013.

3.10.5.4 Exam papers. O Level (see Appendix C for sample English exam paper) comprises of 2 papers which include 2 sections each. *Paper 1* assesses the writing skills of the candidates while *Paper 2* the reading skills.

Time provided to complete *Paper 1* is 1 hour 30 minutes and the marks allocated are 60. The two sections of the paper have different tasks e.g. *Section 1* is Directed Writing in which the candidates are asked to write 200-300 words letters, speech, article or report related to the world of academia, work and community in general. The marks allocated to this section are 30 in which 15 marks can be awarded for task fulfillment and 15 marks for the language. *Section 2* of this paper is Creative Writing in which the candidates are asked to choose one question out of 5 comprising of narrative, descriptive or argumentative essay titles. The word limit for the essay is 350-500 words.

Paper 2 of O Level assesses the reading skills of the candidates and has 2 sections. The texts may be factual or literary and may be taken from a wide range of media or styles. The time allowed to complete this paper is 1 hour 45 minutes and the marks allocated are 50. *Section 1* the candidates are required to scan a factual communication(s) e.g. report(s), article(s), advertisement(s), email(s), or letter(s) of approximately 700 words. The candidates are supposed to identify and write the information about the similarities/differences, causes/effects, advantages/disadvantages, problems/solutions or actions/consequences. Additionally the candidates are also asked to write a summary of about 160 words. *Section 2* comprising of 25 marks in which the candidates read for meaning. They read a narrative passage like a report, article or a story of approx. 700 words and answer short questions which assess both the explicit and implicit meaning. The questions in this section may range from straightforward factual responses on single points to ones requiring a 1 understanding of

the whole text. These usually move from accessible questions to more demanding ones, requiring inferential skills. The topics of texts are not known in advance and thus are unseen.

3.10.5.5 Paper setting of O Level. For O Level, according to Newton, Baird, Goldstein, Patrick and Tymms (2007), the examination papers are developed along with their marking schemes around two years prior to their administration. These papers are designed by the free lance Principal examiners who are experienced examiners and have had some formal training and apprenticeship with the associated board. The exam content is based on the curriculum testing the skills, knowledge and understanding of the candidates keeping in view all the six levels of Bloom's Cognitive Domain in an appropriate proportion (see chapter 2 section 2.3.5 for Bloom's Taxonomy).

3.10.5.6 Marking of O Level. For O Level marking, according to Newton, Baird, Goldstein, Patrick and Tymms (2007), examiners and markers are employed by the boards under temporary contracts. These examiners include practicing or often retired teachers exhibiting considerable understanding of the subject area. The papers or scripts are marked according to the specific marking criteria provided by the board. The papers are sent to the boards from the examination centers from where they are sent to the examiners to be marked. After marking they are returned to the boards and are not sent back to the schools unless there is a formal request made.

Examiners for each subject are structured hierarchically as is the case of FBISE for Matric which be discussed in the latter part of this chapter. Chief examiner who is responsible for the whole examination is at the top, the Chief Examiner reports to a Chair of Examiners who are responsible to maintain a standard across different syllabuses within a single subject area. These Chair of Examiners then report to the highest link in the chain, the awarding body i.e. Accountable Officer who makes the final decision.

The hierarchy a level down of Chief Examiner is a number of Principal Examiners and Moderators who are responsible for the designing of exam question papers and their marking schemes. Under the Principal Examiners is a team of Leaders who are responsible for the training of many Assistant Examiners and Moderators.

Before the marking, a series of training and standardization take place. The more senior examiners train the junior team members to ensure that the Principal Examiner's application of marking criteria is applied uniformly. For the reliability measure, during the marking random samples of papers from different examiners are double checked by the supervisor. If there are inconsistencies in the marking the examiners are advised accordingly and some examiners with too much discrepant marking might not be allowed to continue.

O Level allows the liberty to challenge the awarded grade and request for reassessment or remarking of papers if the schools believe that a student has not been awarded a fair and correct grade. The request for reassessment which is usually known as *Enquires About Results* should be made within a specific period of time. After the request made, the board reassesses the paper and might give the candidate an improved mark to the candidate, or it might remain the same and in some few cases the grade might be lowered. Even after the result of the enquiry, the school still has the liberty to challenge it and lodge an Appeal to the board. Even after the school is discontented with the outcome of the appeal, it might reach out to an independent body i.e. the Examinations Appeals Board.

3.10.5.7 Variability of Examiners/Markers. The central point of the story of testing procedure is the outcome of test scores and the major issue related to it is the variability of test scores. Markers reliability has been the focal point of the research of 19th century and Tattersall's (2007) findings remain relevant to the 21st century. Variability in test scores is extensively dependant on the rater factor (Lumley and McNamara, 1995). The

problem of awarding a reliable mark has been in the focus of the educationists who have drawn attention to this important aspect of assessment. One of the first instances was in 1888 when Professor F.Y. Edgeworth of Oxford in his statistical analysis of university and civil examinations observed that one third of the scripts marked by different examiners received a different mark (Edgeworth, 1888 cited in Tattersall, 2007). Furthermore, one seventh of the scripts received an even different mark when reexamined by the same examiners/markers.

Thus, the reliability of the markers had been questioned for centuries now. Improving the reliability of markers had especially been a focal point of International Conference on Examination in Eastbourne in 1931 and considerable research has undergone since then e.g. as Lumley and McNamara (1995) lists Diederich, French and Calton (1961), Lumley and McNamara (1995), Cason and Cason (1984) works which all suggest that the raters scoring can be as varied as the examinee ability.

So, O Level also addresses this issue by regularly monitoring the marking schemes, examiners and the procedures involved in marking. However, even then where subjective marking is required there cannot be 100% accuracy but there are measures to deal with that e.g. estimating marker reliability by the statistical measures for inter-marker correlations and O Level being an international exam, adapt such procedures to ensure standardized marking (CIE, 2013).

3.10.6 O Level: Validation

There had been many ways proposed in the language testing literature to evaluate the validity of a test. One of the most robust is the Alderson and Buck's (1993) which poses important questions on tests questioning their validity. The criteria proposed by them is useful against which the validity of a test can be evaluated. Below, in Table 3 is the summary

of the desired test validation processes and aspects proposed by Alderson and Buck (1993) adapted into a table by Hawkey, (2004:31).

Test Validation Processes		Aspects and Key Questions
Construct and Content	Syllabus definition	Information on content of exam, purpose, target candidates, difficulty level, typical performances at each grade: information accessible to students? systematic needs analyses of key stakeholders? additional information to item writers?
	Exam construction	Item writers, item writing, moderating, pre-testing: status and training of item writers? Editing and checking? Statistical analyses of pre-tests?
Concurrent and predictive validity		Test validity studies for equivalence of versions and forms: quantitative or qualitative? relationship to awards processes?
Reliability	Administration	Responsibility for administering the exams: training, monitoring?
	Marking	Markers, standardisation, rater reliability and consistency, grade-awards: statistics? double marking? consistency of results?
Construct and Content	Post hoc Analysis and Reporting	Statistical analyses of exams; exam reports and accessibility

	Revision	Exam feedback, systematic revision procedures: student feedback? rationale and frequency of exam revisions?
--	----------	---

Table 3: Summary of the desired test validation processes and aspects proposed by Alderson and Buck (1993, adapted from Hawkey, 2004:31)

Considering Table 3 above for the benchmark of the validation process, O Level satisfies the procedures e.g. the construct, content, concurrent and predictive validity. As regards with the reliability, the administration and marking of exam is quite transparent and the information related to it is easily accessible through internet and published resources.

3.10.7 CEFR and O Level

As, O Level grades can be translated into the levels of CEFR with the comparability grid for comparing the achievement on O Level, so it can be stated with confidence that it is aligned with the CEFR. Many studies of validation and reliability are also provided which makes O Level a world recognized test as preferred over the local testing system.

O Level grades are translated in terms of levels of CEFR as can be seen below in the Table 4. A* are not awarded at component level and only in the aggregate. Even at the individual component level the ceiling is B2 in reading and C1 in writing and yet it is possible to attain C2 in the aggregate if one gets A*. So the difficulty level is quite high for O Level papers. However the equivalency Table 4 below is for IGCSE because many of the schools are adapting IGCSE. However, O Level and IGCSE are the products of the same testing bodies and hold the same weightage according to the official website of Cambridge CIE. The grade structure is, however, slightly different only in the number of grades awarded, for example, O Level does not have F and G for the lower achieving students as GCSE or IGCSE have.

Component level	Reading	Writing	Listening	Speaking (Count-in, syllabus 0511)		Speaking (Separate endorsement, syllabus 0510)
Grade A	B2	B2-C1	B2	C1	Grade 1	C1
Grade B	B2	B2	B1	B2-C1	Grade 2	B2-C1
Grade C	B2	B2	B1	B2	Grade 3	B1
Grade D	B2	B2	B1	B1	Grade 4	A2-B1
Grade E	B1	B1	A2	B1	Grade 5	A2-B1
Grade F	A2-B1	A2	A2	A2-B1		
Grade G	A2	A2	A2	A2-B1		

A* not awarded at component level

Table 4: CEFR levels of Cambridge IGCSE English as a Second Language (E2L) by skill? adapted from the official website of Cambridge CIE³⁶

3.11 Matric

Matric, a short form of matriculation, refers to the final examinations of class 9 and 10 of secondary school. It is a school leaving examination in Pakistan after which the student is awarded Secondary School Certificate (SSC). It is a pre-requisite to get into the college to proceed to the intermediate level (an equivalent of A Level) on the completion of which Higher Secondary School Certificate (HSSC) is awarded to the student. It is considered as an equivalent of O Level, UK GCSE, or its international version IGCSE by the IBCC in Pakistan.

³⁶Retrieved from
http://ask.cie.org.uk/system/selfservice.controller?CONFIGURATION=1035&PARTITION_ID=1&TIMEZONE_OFFSET=&USERTYPE=1&CMD=VIEW_ARTICLE&ARTICLE_ID=534926

3.11.1 History of Matric

Matriculation, according to the Merriam-Webster dictionary means, “to enroll as a member of a body or especially of a college or university” derived from the Latin word *matricula* meaning “public role”. It is, however, used to refer to getting enrolled in the college or university as an academic requirement in many parts of the world e.g. in universities of Oxford, Cambridge etc. Matric, in many commonwealth countries refer to successful completion of a secondary school examination e.g. India, South Africa, Bangladesh and Pakistan etc. Matric is based on British educational system. A more Westernised education system was introduced in the British India by the Viceroy of British India, Lord Macaulay then (Gupta, 2007). The purpose of introducing Westernised education system in the British colonies, according to (Punchihetti, 1993 cited in Wall, 2005), was to create a lower supportive administration work force and more importantly it was also an attempt to transfer value system of the more dominant culture. And along with all these underlying motives English started to become the core subject of the curricula.

3.11.2 Use of Matric in Pakistan

Matric is administered by different educational boards (see section 3.3 on IBCC and the Educational Boards of Pakistan above) through extensive networks of centres throughout Pakistan. It provides a range of subjects classifying the candidates into two main groups i.e. Science and Humanities. In comparison with any secondary school international qualification e.g. O Level, GCSE, IGCSE, etc., Matric is chosen by most of the student base in Pakistan. The reason of its popularity is because of its low cost and because it is readily available in many parts of the country.

3.11.3 Key Numbers of Matric

Matric is taken by a significant number of students each year. According to the latest figures provided by the FBISE officials in an interview with the researcher, 64327 students took SSC in the year 2012 conducted by only one board i.e. FBISE³⁷. Similarly, the 8 Boards of Intermediate and Secondary Education (BISE) in Punjab 206,469 sat for Matric in 2012 (The Nation, 2013)³⁸.

3.11.4 Curriculum for English: Matric

The National curriculum for English Language (2006:2) by Ministry of Education, Pakistan claims:

“....the curriculum places greater emphasis on the understanding and use of the English language in different academic and social contexts than on acquiring knowledge about the language for its own sake. Such an approach acknowledges, on one hand, the importance of teaching the knowledge about the language system; on the other, it moves a step forward to emphasize the appropriate use of that knowledge so that students' ability to communicate in real life situations is improved and made effective for various purposes.”

The curriculum then undoubtedly is based on the communicative approach. According to Council of Europe (2001), the pedagogic tasks in a language classroom where communicative approach to language teaching is employed are indirectly related to real life situations. The main aim is to develop the communicative competence of the students. These authentic tasks promote meaningful communications which are relevant and challenging, they promote critical thinking of the pupil and thus lead to tangible outcomes. The Matric English curriculum as can be seen in Appendix A are clearly rooted in the communicative approach to teaching.

³⁷ Information provided by FBISE officer in an interview on 17 December, 2012.

³⁸ Information retrieved from nation.com.pk on 17 June, 2013.

The National curriculum for English Language (2006) by Ministry of Education, Pakistan seems very promising with the aim to reform the textbook writing and selection procedures within the educational boards and more specifically aiming to guide the boards towards a better standard of English teaching and testing standards.

Although, IBCC under Ministry of Education and Training is responsible for examination reforms, implementation of curriculum and quality assurance across the boards but how far is it all practiced is a matter of investigation. It claims to ensure coordination of educational activities like curriculum, testing and evaluation, examinations, educational planning and administration promoting centralized education but since the ultimate authority of implementation of curriculum and quality assurance is with the provincial governments and the boards of secondary and higher secondary education at the provincial level and because of that the standard varies from province to province as well as from board to board. The inconsistency then lies within the implementation of this curriculum which if the question paper is studied does not adhere to the standards of CALTe.

3.11.5 Matric Examinations: Approach and Specifications

The case of Matric is similar to Sri Lankan O Level which was evaluated by Wall (2005) in the 'Sri Lankan O Level Evaluation Project'. Matric English exam also does not follow CALTe as the test content is mostly based on the text book taught in the class. The question paper includes mostly questions from the text book requiring candidates to use their memory and not communicative competence, creativity, critical skills. For example, the very first MCQ of the Model paper available at the official website of FBISE is:

"Lok-Melas are arranged to develop a sense of A) selfishness, B) Brotherhood, C) Disharmony, D) Chaos." (FBISE, 2013)³⁹.

³⁹ retrieved from www.fbise.edu.pk on June 12, 2013

There is a serious reservation about such type of questions in a reading comprehension paper. How is it possible to answer such a question without any reading passage given in the question paper? Therefore the response is prepared in advance. There are many such like questions in section 1 which require memory than any reading or creativity skills. While exploring the impact of testing, Vernon (1956) one of the first educators to talk of measurement of mental abilities, claimed that examinations promoting cramming and rote memorization could stunt the reasoning power of the pupil.

Again, the very first question of the section 2 is:

“What teaching methodology did the Prophet Muhammad (SAW) use to teach the Muslims?”

Again, without any reading passage, it can be safely deduced that it requires the candidates to reproduce something that they have memorized previously. It is also taken from a lesson from the textbook “Hazrat Muhammad (SAW) an Embodiment of Justice” (Unit 1, English Text book for class 10). Similarly, the whole section 1 comprises of the question based on the textbook lessons.

However, the last part out of three which is a Writing test does incorporate a few questions which might demand the candidates to use their communicative competence and higher order skills of Bloom’s Taxonomy. Nevertheless, the same paper pattern for English is followed year after year which makes the test content very predictable. Generally, the teachers and students know the paper content before hand as it is based on the text books and prepare for the exam accordingly. During many personal interactions with the students, it was revealed that they also memorized whole essays for the Writing section of the paper as each year similar topics are given to write an essay or letter about.

Matric is a combination of objective and subjective assessment but more objective in nature. Matric being more objective in nature can be because of the fact that it is more inclined towards the psychometric-structuralist approach than the communicative approach.

In regards to the specifications for Matric, they are nevertheless, chalked out to some extent giving some information on the test content, the length and structure of the test tasks, the type of material used whether authentic or unauthentic only in the form of sample questions provided in the sample question paper. The information on the type of response expected of the candidate and the scoring methods and criteria are, however, missing⁴⁰.

3.11.5.1 Exams and Marking Scheme of Matric English Papers. Matric is exam-based and a summative test. Its English test is designed to assess many of the same language abilities as O Level; reading and writing. Being an achievement test it is based on a prescribed syllabus. English is one of the core subjects and is a compulsory component of the examinations for both the major groups i.e. Science and Humanities. Matric is generally taken in 7 subjects across the curriculum after 11 years of education (from class prep to class 10).

The exams for Matric are conducted annually in March/April every year. Matric also conducts a supplementary exam after a couple of months of the annual exam for the students who want to re-sit the exam either because of failing in the previous attempt or if the candidate wants to improve their score.

Matric does not comprise of the grading system like O Level but instead numerical scores are awarded. A candidate securing less than 33% marks is considered fail and is required to re-sit the exam of that specific subject. The equivalence formula by IBCC converts the grades of O Level into numerical scores according to specific criteria (see

⁴⁰ See www.fbise.edu.pk

Appendix E). The accreditation and equivalence is done for the entrance tests of colleges and universities.

Besides the detailed grade descriptor available for each grade of O Level, O Level English exam can be further interpreted as an evidence of the candidates proficiency in English language as they are aligned to the scales of CEFR which clearly states information about the linguistic ability of the candidate with the help of comprehensive descriptors. While the numerical scores of Matric does not provide any evidence about the communicative language ability of the candidates.

The marking key provided to the examiners of Matric relies basically on holistic marking⁴¹. Furthermore, Matric students are only given their English marks without giving any information about what a 90% or 40% marks achieving candidate can linguistically perform in English. Thus, the numerical score does not tell anything what a testee is capable of performing linguistically in English at 40% or 90%.

3.11.5.2 Matric English question papers. Language is needed to survive in the social environment and communicate effectively and efficiently. The main objective of learning a foreign language is to be able to communicate efficiently in all the concerned domains whether public, occupational, educational or personal. The indispensable and the ultimate aim of language learning is to develop the learner's linguistic competences (Council of Europe, 2001). The aim of language teaching should be to enable an individual to mobilize his/her linguistic competences in order to implement and possibly improve or extend them as has been seen the reason of paradigm shift of teaching, learning and assessment. The question arises then how can memorizing certain answers to a possible set of questions from the text book help in language learning? which is a persistent trend in Matric English question papers. The Matric English paper consists of 3 sections. The first 2 sections are

⁴¹ See chapter 2 for definition

mostly based on the prescribed text book and a candidate with good memory can easily sail through the 2 sections yielding a passing mark i.e. 33%. There certainly is a role of memorising language chunks like phrasal verbs, idioms, and vocabulary in general but to be able to use them in real life situations needs competence and skills to be developed. As Bachman's (1990) model of Communicative Language Ability (CLA) suggests the development of the learners' language knowledge, strategic competence and psychological mechanism.

3.11.5.3 Course objectives and test items of Matric. The English curriculum of Matric as mentioned above and its objectives (see Appendix A) clearly categorizes it a communicative syllabus but the test items where the memory skills are tested negate the existence of any communicative goal. For example, in *Section A* of Matric English paper whose weighting is 20% consists of Multiple Choice Questions (MCQ) which are mostly taken from the syllabus text book. Similarly, *Section B* whose weighting is 50% comprises of short answers to questions. Out of the 50% weighting almost 25% is again are mostly based on the text book. So, a test where the pass mark is 33%, if $20\% + 25\% = 45\%$ can be acquired by simply memorizing answers to predictable questions, the validity of such a test is seriously questionable. As Morrow (1977) identifies unpredictability of test items/tasks to be one of the crucial components of a good quality test and if this quality is lacking can put the validity of such a test in serious doubt. So, in such a case where almost half of the test could be predictable raises questions.

3.11.5.4 English exam papers of Matric. The English exam for Matric English is taken in two parts in two years of the secondary school i.e. the candidates get an aggregate of SSCI and SSCII (class 9 and 10). So, in a way, Matric students are at an advantage to have

less burden than the O Level students who appear for their English test in one go. So, in theory Matric students can therefore prepare better.

Matric requires the candidates to take English Compulsory exam in both the years i.e. Secondary School Certificate (SSC-I) and SSC-II which are equivalent to year 10 and year 11 of UK education system. SSC-I and SSC-II (see Appendix D) has 2 papers comprising of 3 sections each. Every year there is 1 paper comprising of both reading and writing skills, unlike O Level where the students take 2 papers separately for each skill and of shorter duration. The time provided to complete each paper of Matric is 3 hours and marks allocated for each paper is 75. Marks from SSS-I and II are combined and the testees get the aggregate marks from both the years at the end. Each paper has objective and subjective sections. The language sub-skills that are being assessed are spelling, reported speech, voice transformation, synonyms, gap-filling, grammar, using phrasal verbs in sentences etc.

Section A (weighting 20%)⁴² consists of Multiple Choice Questions (MCQ) which has to be attempted within a prescribed time usually 20 minutes. The MCQs are mostly extracted from the syllabus text book.

In **Section B** (weighting 50%) candidates are required to answer short answers which are again mostly based on the text book, so rote learning could come very handy here. There are also some extracts of either poetry or prose or sometimes both which are taken from the text books and the candidates are required to comprehend and answer accordingly. If the candidate has already learnt the answers, he might even answer the questions from his memory without even reading the text. With question types like completion of textual references, it seems as memory is being tested rather than the language skills.

In **Section C** (weighting 30%) candidates write long essay type questions which are usually letters, dialogues, essays, and applications. In SSC-I there is also a question relating

⁴² See Appendix D. English syllabus and sample papers can also be accessed on www.fbise.edu.pk

to translation where the candidates are required to translate a few sentences from Urdu to English.

Section A and the first question of **Section B** are mostly based on the text and require mostly memory skills of the candidate. The weighting of both the sections is 70%, out of which 45% can generally be yielded with good memory skills.

3.11.5.5 Paper setting and marking of Matric English papers. In an interview with some of the staff members of FBISE, the following information was given about how English papers are set and marked.

3.11.5.6 English Paper setting of Matric. FBISE has established an Academic Committee of the Board which includes representative from different institutions and directorates from different districts. The academic committee includes: the Chairman of the Board, the Director General, Federal Directorate of education, Islamabad, the Director of Education, FGEI (C&G), GHQ, Rawalpindi, the Director of Education, Northern Areas, Gilgit, the Director of Education, PAF Rear Air Headquarters, Peshawar, the Director of Education, DHA, Karachi, the Director of Army Education, GHQ, Rawalpindi, a nominee of the University Grants Commission, a nominee of the Ministry of Education, two nominees of the Federal Board, a representative of the Curriculum Wing, Ministry of Education, two Principals of College of whom one shall be a woman, one Headmaster and one Headmistress of two schools, five conveners of the Committee of Courses, a representative of Islamabad Model Institutions and a representative of institutions under Board of Governors.

This Academic Committee nominates and approves subject specialists who are primarily teachers and have at least 10 years of teaching experience. Set guidelines, syllabi, paper patterns and forms to be filled are sent to them for paper setting. Once the papers are set by the paper setters, FBISE under the supervision of another specialist committee makes

sure the papers are within the syllabi. The subject specialists are trained by trainers (who are inducted by the Director Research and his team of senior specialists) if there is any change in the paper pattern. In an interview with the researcher, FBISE staff informed that no such training of subject specialists and paper setters takes place.

3.11.5.7 Marking of Matric English papers. Based on the information received in the interviews conducted with the FBISE officers in October, 2010 and December, 2012, FBISE selects the head and sub examiners/markers through requesting different educational institutions to nominate them. A great deal of importance is given to experience. The sub examiners have 5 years of teaching experience while the head examiners have 10 years experience. Before the marking of papers, all the head and sub examiners have "standardizing" meetings to discuss the Question papers and make sure that nothing is out of the syllabus (or in other words nothing is out of text book). They prepare an answer key together which is subsequently distributed amongst all the examiners/markers. The sub examiners mark all the papers out of which randomly selected 10 papers are double marked by the head examiners for the reliability of the marking.

However, there is no training for the examiners/ markers to make sure the standardization as should be the case where examiners are trained through marking sample scripts and assessment criteria discussed thoroughly with other examiners and head examiners. Discussion of the assessment criteria is done to some extent but the assessment criteria itself is not very thorough⁴³ as are the O Level's ones.

There is no tradition of estimating marker reliability by the statistical measures for intermarker correlations which should be a routine practice for any high-stakes public exam to ensure a high standard.

⁴³ Information received in an interview with FBISE officer in October, 2010 and December 2013.

SE does not allow reassessment or remarking of papers. It is basically recalculations of the marks for each question which is permissible. So, there is no liberty to challenge the awarded grade and request for reassessment if the student believes that he/she has not been awarded a fair and correct grade. There is no measure for any appeals as in O Level to reach out to an independent body i.e. the Examinations Appeals Board.

Interviews with the members of FBISE (October, 2010 and December, 2013) made clear that nothing of the sort happens in Pakistan due to the fear of opening a Pandora box that might not be handled appropriately.

In all, the need for Matric is to follow such a mechanism which has already been developed by CEFR like O Level which is an internationally recognised test and deals with many important aspects of the examination process very thoroughly.

3.1. Variability of Examiners/Markers. Keeping in view as discussed in the previous chapter, the history of reliability of raters starting from Professor Edgeworth in 1888 till the present time, this issue is addressed at variable levels in different testing systems. This issue is not only raised in Pakistan but is a worldwide phenomenon needing consideration.

variability and reliability of raters can be reduced with regular trainings. More detailed marking scheme is yet another possibility to achieve reliable marking which was realized as an important step for the reliability of scores as early as 1929 when Ruch (cited in Linacre, 1991) argued that subjectivity of marking can be reduced by as much as about one-half if there is conformity to a set of marking rules especially when essays are being marked. As a solution to the problem of reliability of scores, a more traditional technique of double or multiple marking is considered as another possible answer but limitation of resources often makes it impossible to bear the expenses of more examiners.

Matric English test, however, employs many examiners working under the supervision of a head examiner. There is, however, no trend of estimating markers' reliability through statistical measures.

3.11.6 Validation of Matric English Exam

Validity being the most important virtue of any good test is a concern for the test users and providers worldwide. There are different criteria proposed by many. The desired test validation processes and aspects proposed by Alderson and Buck (1993) adapted into a table by Haynes, (2004:31, see Table 3 above) are seriously lacking in Matric. From the interviews held with FBISE officer in October, 2010 and December, 2013, it can be concluded that Matric does not provide any satisfactory proof of validity, for example neither *construct content, concurrent*⁴⁴ nor *predictive*⁴⁵. As Kellaghan and Greaney (1992) also suggest in order to improve the examination system, each of the examination board should have the capacity to conduct research and that predictive validity studies of public examinations specifically should be conducted. They also draw attention to the need of development of professional competence of examination authorities in test construction and that educational planners, examination authorities and curriculum developers should work closely together.

As Matric English test is more of a memory test as seen above, so the content validity resembles there. As for the concurrent and predictive validity which evaluates the equivalence of different versions of the test either qualitatively or quantitatively there is no evidence provided. There is no training for the item writers, no pre-tests so consequently no statistical analyses for that. Reliability of markers and their consistency is dealt with to some

⁴⁴ According to Hughes (1989) concurrent validity means when the criterion of the test which is supposed to be a highly dependent assessment and the actual test are administered at the same time and produce the same results. (see Hughes, 1989:23 for detailed discussion).

⁴⁵ See footnote 44 for definition

extent (section 3.11.5.7). Statistical analyses or any reports related to the validity of Matric English test are not available. Revision of test content and its overall procedures is yet another neglected aspect. Although, the exams were tried to be revised with the help from Edexcel exam in (2005) but was not fruitful and slowly and gradually FBISE is reverting back to the old ways of exam based on rote learning. The reason of its failure was because only FSC tried to design a more skill based test while the other educational boards still preferred the memory based tests. Therefore, the relevant testees found the skill based test difficult which was evident from the decline in the success rate of the testees.

3.11.7 Problem: R and Matric

Even into account the qualitative analysis of Matric question papers and the testing procedures raises serious doubts about the equivalence of both the secondary school assessment systems. It is not being cynical or a mere desire to deliberately portray Matric below standard but even from the evidence above one would be skeptical and raise questions about the nature of Matric English examinations. It might be deduced that Matric English test does not comply with CEFR. To further prove the point, the alignment to CEFR in the phase 2 of the present study and some quantitative analysis in the first phase of the study was also carried out to compare both the systems.

3.11.8 Problem: Conversion formula of IBCC

Established fact of language certification issues is one that of transparency. In Pakistan this issue concerning a high-stakes exam such as O Level is very pronounced. Many students opt Matric only because of the discrepancy of the IBCC scale or formula (see Appendix E) to convert the scores of students with foreign qualifications into

⁴⁶ As per the conversion formula (IBCC official website, 2013) students who secured A* can be granted equivalence up to 90 percent. Grades will be equal to 85 percent, Bs will get 75 percent, Cs will be awarded 65 percent, 55 percent will be given against Ds, while Es will be equal to 45 percent.

scores corresponding those of the local grading system. IBCC determines the accreditation and equivalence to foreign qualifications with corresponding Pakistani Certificates at the Secondary School Certificate (SSC, Matriculation), Higher Secondary School Certificate (HSSC) (Intermediate levels) and technical education as well as local certificates/diplomas with the comparable certificates awarded by the educational boards. It is like the Australian Average Marks Scaling which converts or rescales the scores into a different currency to make them comparable to a specific exam and for use in other functions (Coe, 2007). The average cases of equivalency are around 15000 per year which are processed by IBCC. But the students usually do not agree with it and believe that they are given fewer marks than they actually deserve after the conversion of grades into numerical scores and therefore are not the true representation of their English language abilities. As per IBCC's conversion scale, a student receiving A* cannot be granted more than 90% even if his raw score is 96 in O Level. Consequently, many high achievers of O Level are barred from getting into professional colleges and universities which require a very high local score. Thus, many such students are forced to opt for Matric and are, therefore, deprived of any other system that they might desire to choose. Interpretation of certificates needs a common, fair and transparent basis, although this study is only concerned with the accreditation of English language but the whole system needs to be revamped in order to provide fair and just.

CHAPTER 4

RESEARCH DESIGN FOR THE STUDY

This chapter explains the research methodology adapted for this research study. At first, it introduces phase 1 of the study. It explains the procedures and materials employed for this phase of the research. The latter part of the chapter it is concerned with the phase 2 of the study; procedures and materials for phase 2. It explains the methodology adapted for aligning Matric to CEFR followed by the results yielded.

4.1 Research Methodology

This chapter will discuss the methodological derivation especially drawn for this study. Research hypotheses and the terms are discussed in detail along with the research methodology and tools adopted. The study was classified into two phases; Phase 1 dealing with the quantitative strand and Phase 2 with the qualitative. The first phase, focused on the quantitative part of the study where the Common Test Approach was adopted in order to compare the scores of students from Matric and Level in English. In the second phase, on the other hand an attempt was made to align Matric English test/exam (administered by FBISE) to CEFR. All the test content, test procedures and practices were carefully analysed through the linking process suggested in the CEFR Manual (Council of Europe, 2009, henceforth the Manual).

CHAPTER 4

RESEARCH DESIGN FOR THE STUDY

This chapter explains the research methodology adapted for this research study. At first, it introduces phase 1 of the study. It explains the procedures and materials employed for this phase of the research. The latter part of the chapter it is concerned with the phase 2 of the study; procedures and materials for phase 2. It explains the methodology adapted for aligning Matric to CEFR followed by the results yielded.

4.1 Research Methodology

This chapter will discuss the methodological derivation especially drawn for this study. Research hypotheses and the terms are discussed in detail along with the research methodology and tools adopted. The study was classified into two phases; Phase 1 dealing with the quantitative strand and Phase 2 with the qualitative. The first phase, focused on the quantitative part of the study where the Common Test Approach was adopted in order to compare the scores of students from Matric and Level in English. In the second phase, on the other hand an attempt was made to align Matric English test/exam (administered by FBISE) to CEFR. All the test content, test procedures and practices were carefully analysed through the linking process suggested in the CEFR Manual (Council of Europe, 2009, henceforth the Manual).

Although, the purpose of the study was to compare English tests of Matric and O Level in Pakistan, the main emphasis was, however, laid on Matric, especially in the second phase of the study. Furthermore, the comparability of the official test scores from both the tests was also investigated in the first phase of the study.

The first four hypotheses dealt with the quantitative strand of the study. While the last two, more qualitative in nature, related to the alignment of Matric to CEFR were attempted to be tested in the second phase of the study. It was qualitative in nature because it was the content analysis of Matric English exams including the exams procedures, papers, etc. Below is the review of the hypotheses along with the operational definitions of the key terms derived for this study.

Hypotheses 1: The conversion scale provided by IBCC of converting scores from O Level to Matric underwrites the performance of O Level students by comparison with Matric students.

Conversion scale is operationally defined as the formula developed to convert scores from one examination system to another in order to find an equivalence scale. This study tried to find out whether:

- the O Level students are given the score that they deserve when their O Level grades are translated into Matric numerical scores?
- the conversion scale from O Level to Matric is appropriate and fair?

The conversion scale provided by Inter Board Committee of Chairmen (IBCC, see Appendix E) was probed into to see whether it undervalues the performance of O Level students while converting their scores from O Level to Matric in comparison with Matric students.

• the O Level students are given the score that they deserve when their O Level grades are translated into Matric numerical scores?

• the conversion scale from O Level to Matric is appropriate and fair?

tried to find out whether:

Conversion scale is operationally defined as the formula developed to convert scores from one examination system to another in order to find an equivalence scale. This study tried to find out whether:

- the O Level students are given the score that they deserve when their O Level grades are translated into Matric numerical scores?
- the conversion scale from O Level to Matric is appropriate and fair?

tried to find out whether:

Hypotheses 1: The conversion scale provided by IBCC of converting scores from O Level to Matric underwrites the performance of O Level students by comparison with Matric students.

Conversion scale is operationally defined as the formula developed to convert scores from one examination system to another in order to find an equivalence scale. This study tried to find out whether:

- the O Level students are given the score that they deserve when their O Level grades are translated into Matric numerical scores?
- the conversion scale from O Level to Matric is appropriate and fair?

tried to find out whether:

Below is the review of the hypotheses along with the operational definitions of the key terms derived for this study.

Hypotheses 1: The conversion scale provided by IBCC of converting scores from O Level to Matric underwrites the performance of O Level students by comparison with Matric students.

Conversion scale is operationally defined as the formula developed to convert scores from one examination system to another in order to find an equivalence scale. This study tried to find out whether:

- the O Level students are given the score that they deserve when their O Level grades are translated into Matric numerical scores?
- the conversion scale from O Level to Matric is appropriate and fair?

tried to find out whether:

The first four hypotheses dealt with the quantitative strand of the study. While the last two, more qualitative in nature, related to the alignment of Matric to CEFR were attempted to be tested in the second phase of the study. It was qualitative in nature because it was the content analysis of Matric English exams including the exams procedures, papers, etc. Below is the review of the hypotheses along with the operational definitions of the key terms derived for this study. Furthermore, the comparability of the official test scores from both the tests was also investigated in the first phase of the study.

Although, the purpose of the study was to compare English tests of Matric and O Level in Pakistan, the main emphasis was, however, laid on Matric, especially in the second phase of the study. Furthermore, the comparability of the official test scores from both the tests was also investigated in the first phase of the study.

4.2 Hypotheses

students are sceptical towards the claim that IBC makes through its conversion scale (see Cambridge test i.e. O Level, IGCSE and has an equivalence formula for that. However, many 3.11.8), the scores from English test from Matrix are equivalent to any Secondary School the inequality of both the testing systems. As claimed by IBC (see Chapter 3, section 3.11.8) the aspect of English language i.e. reading and writing. The thesis is based on the assumption of are meant for the same level i.e. Secondary level and are supposed to measure the same Matrix. To see how meaningful both the English tests can be compared, since both of them are meant for the same level i.e. Secondary level and are supposed to measure the same aspect of English language i.e. reading and writing. The thesis is based on the assumption of are meant for the same level i.e. Secondary level and are supposed to measure the same

O Level is not equivalent.

Hypothesis 3: Matrix English testing system is compared to English testing system of O Level
are used in this study to make interpretations about the validity of their secondary test scores. are used in this study to make interpretations about the validity of their secondary test scores. programmes (Bachman, 2004). The test scores from Matrix, O Level and the Common Test knowledge or what he/she can do in a certain language and to make decisions about people or The scores from the language tests are used to make inferences about the candidate's

significantly higher than the Matrix students on the Common Test or vice versa?

equally good or bad on the Common Test? Whether O Level students score better and both the systems have similar English reading and writing skills? Whether they perform primarily trained for these two language skills. The aim was to test whether students from rationale of choosing a subtest of IELTS) because the students in both the systems are test was the reading and writing modules only of IELTS (see section 4.9 below for the to determine the English language competence of both the groups. The Common Test in this Common Test Approach (see below section 4.5, Research Framework) was employed

and writing than Matrix students.

Hypothesis 2: Students of O Levels score better on a Common English Test of reading

Levels and the Common Test (subtest of IELTS) have already been aligned to CEFR, Matric compared their previous official scores with the scores from the Common Test. Since O meaningfully, the researcher using the Common Test Approach (see section 4.5 below) scores from O Level and Matric English Language testing can actually be compared by many which this is trying to prove with substantial evidence. Therefore, to see if the same score in both the testing systems but in practice this might not be the case as argued a way a claim could be made that students with the same linguistic competence are awarded should be treated the same. So, in theory, both the assessment systems are comparable and in As per IBC conversion scale, Matric and O Level are equivalent and therefore

inference from them about the English linguistic abilities of the testees.

are reliable. So, are Matric and O Level English scores reliable enough to deduce any on the other hand, are operationally defined as the scores from which the inferences deduced linguistic abilities as they claim to in the objectives of the English syllabus. Reliable scores, and O Level English scores would be whether they are truly representing the students and meaningful with regards to the purpose of the test (Brown, 2003). The validity of Matric and Harrison, 1983) and whether the inferences drawn from the results of a test are appropriate linguistic which means whether the test is measuring what it is intended to measure Validity, is a common and the most important notion in educational and applied

Hypothesis 4: English test scores from O Level are more valid and reliable than Matric.

both the streams in the first phase of the study. compared in the first phase preceded by the comparison of performance of the students from most recent test papers of last five years (i.e. 2008-2012) from both the systems were was scanned thoroughly through the linking process suggested by CEFR Manual (2009). The Appendix E). To evaluate if both the systems are comparable, Matric English testing system

specification, test design, question papers, marking criteria/ schemes etc. It was more testing procedures evaluated in this study included the process from test CEFR, an international framework, to which O Level already is aligned. same level of validity as O Level when evaluated through the alignment process of Hypothesis 6: **Matric English testing procedures practiced by FBISF do not possess the same linking Manual to CEFR (2009).**

- map the tasks/ questions of English test paper of Matric to procedures outlined in the
- determine the level of CEFR that the Matric English exam tasks are meant for.

Level English paper in that regard.

- determine the level of difficulty of Matric English exam tasks and whether they match O process to CEFR was to:

were analysed in depth. The aim of scanning the Matric's English papers through the linking of previous five years i.e. 2008- 2012 (as was also proposed in the CEFR Manual, 2009) suggests to test the students on. The types of questions and tasks in the Matric English papers English test matches in the difficulty level to the level of difficulty appropriate to the level it the level of difficulty of Matric English test was determined in order to see whether Matric Level being a relative term was gauged through CEFR. Using the levels of CEFR, Cambridge.

Hypothesis 5: **Matric English testing procedures practiced by FBISF are not testing the same level of English reading and writing skills of students as O Level by CIE,**

are compared and linked through the scores of the Common Test.

Council of Europe's Manual (ESOL, 2011). So, the previous scores from Matric and O Level (in this case, IELTS) linked to the CEFR which is also a recommended approach in the can also be linked to CEFR through indirect linkage via 'equation' to an already existing test

frequently. Matric tests, in general, and English papers, in particular, have little or no way of acceptable degree. Cambridge tests are regularly monitored, updated and put to scrutiny design these tests and make sure they possess all the test qualities to a reasonable and been scrutinised equally but since O Level is a Cambridge exam, the world top rank experts⁴⁷ Even though, it was a comparative study and ideally both the tests should have study or work abroad.

CEFR, it will prove beneficial to the Matric users in Pakistan especially the ones who plan to bring in Matric at par to any internationally recognised test. If Matric could be aligned to international framework in general. The aim was to provide positive criticism in the hope of CEFR which will prove its actual comparability to O Level and the standing of Matric in an was attempted to see if the Matric could be aligned with an international framework like only locally but also by the international community. Linking of Matric English test to CEFR levels scores can be translated into the levels of CEFR which are easily interpretable not language examinations are aligned to CEFR, so O Level is also aligned to it and therefore O scanned through an international framework i.e. CEFR. Since, most of the Cambridge Again if the two systems in question are comparable, can they be successfully

- Does all the testing procedures conform to the standards of the Manual?
- Does the test provide proof for its validity and reliability?
- Does the test in question conform to the standards set by the Manual CEFR?

the CEFR. The questions put forth in an alignment process to CEFR include: measure. Alignment is operationally defined as the process which makes a test comparable to study refers to, whether the content included in the exam is testing what the test is meant to qualitative in nature because it was content analysis of the whole system. Validity in this

CEFR to investigate the overall quality of Matrix English test. The second phase of the study, on the other hand, was delimited to examination standards). The second phase of the study, on the other hand, was delimited to detailed discussion on the techniques available for monitoring the comparability of available for such a task (see Newton, Baird, Goldstein, Patrick & Tyms, 2007 for a detailed discussion on the techniques available for monitoring the comparability of available to compare the tests in question was Common Test Approach out of many options of varying reasons. In order to be focused and lead to tangible outcomes, the approach selected to fulfill their promises, the rest either did not respond at all or refused to help because study out of which 9 schools agreed to take part in the research but eventually only 5 were or whether they offered both Matrix and O Level (as many private schools only offer O Level in Islamabad). Initially, all the private schools were considered for the inclusion in the because schools were selected according to whether they were from private or public sector, included. The sample was a „a stratified purposeful sample“ (Miles & Huberman, 1994: 28) these factors. Therefore, for the first phase, only 5 private schools in Islamabad were geographic and socioeconomic conditions as language is thought to be most influenced by To have a homogeneous group, it was essential to select a group with similar

4.3 Delimitations of the Study

papers was deemed important rather than the O Level. papers are based on an international framework, the content analysis of Matrix English available for the O Level English testing procedures ensuring the fact that O Level English methods are used. So, considering all the easily accessible reliability and validity evidence also not ensured, even the familiarity of these terms is also not common and no psychometric ensuring their validity formally i.e. do they test what they are supposed to test. Reliability is

in this research study (see section 4.6 below), which helped in comparing both the Quantitative (Phase 1 of the study) and Qualitative (Phase 2) methods were integrated

4.5 Research Framework

analysed.

In January 2013, all the data was encrypted in MS Excel sheets ready to be statistically consent from one of the examiners, an effort was made to align Matrix English test to CEFR. mark them (see the section 4.8 on Raters below). During the same time, with the help and writing scripts were subsequently distributed amongst Cambridge certified examiners to write. In November-December 2012, students were assessed on a Common Test and the Matrix. and Department of Theoretical and Applied Linguistics (DTAL, Cambridge, UK) it was agreed upon to use CEFR for the content analysis and testing practices and procedures of research methodology with the supervisor in Pakistan and experts in Cambridge Assessment contacted in Summer of 2012, many of who promised to help. After deciding upon the information about IELTS and O Level exams. Different schools of Islamabad, Pakistan were conducted from time to time with AEO and British Council, Islamabad to acquire 18th, 20th and 21st December 2012 respectively. Moreover, informal interviews were also Third, fourth, fifth and sixth were conducted with different members of the board on the 17th, the following order; second meeting on the 14th of December 2012 with the Chairperson, was conducted in October 2010 which was very informative and the basis for future collaboration between the researcher and the board. The subsequent interviews were held in FBISE during the course of two years. The formal ones were 6 in number. The very first one Many formal and informal interviews were conducted with the senior officials of

4.4 Time Framework for the Study

boards then (Willmott, 1980 cited in Murphy, 2007). From 1965 to 1985, as Murphy (2007) (CSE) was introduced. It played its role by providing guidance to the newly established CSE application can be traced back to 1960s in the UK when Certificate of Secondary Education world's leading expert organisations in the UK, Australia, Europe, USA etc. Its early defend it by arguing that if it were so flawed it would not have been able to survive in the own limitations (Murphy, 2007). The proponents of this approach, as Murphy (2007) notes, Tymms, 2007) and so does the Common Test Method, which being one of the list, has its methods have their strengths and weaknesses (see Newton, Baird, Goldstein, Patrick & at the same time has also attracted controversies over the years. Since, all the comparability has been appreciated by its proponents and is considered as one of its important strengths but Although, this approach is fairly easy to understand and quite straightforward, which this case are Matic and O Level.

sample of students in order to compare the standards of two exams/tests in question, which in years of different subjects. In this method an additional Common Test is given to the same of students who have taken an examination through different awarding bodies or in different providing substantial results (Murphy, 2007). This approach is used for comparing a sample approach has been included for many years now and has proved its worth over the years by two boards, subjects, years, etc. (Murphy, 2007). In the comparability studies, Common Test examination. Common Test Approach is usually used for comparing the standards between outlined in its Manual (Council of Europe, 2009) for the quality analysis of Matic English scores of Matic and O Level, while phase 2 uses CEFR and the materials and procedures Phase 1 uses the Common Test Method for the purpose of comparability of the test the respective groups of students.

assessment systems i.e. O Level and Matic and for assessing the English language skills of

statistical procedures. While, on the other hand, using qualitative methods the exam scores were tested by comparing them with their scores of the Common Test through using quantitative methods the validity and reliability of the candidates' previous official goals and circumstances of the study and to come to a significant conclusion. For example, Both quantitative and qualitative research methods were combined according to the

4.6 Rationale of Using Qualitative and Quantitative Methods

educational sectors involved in the educational process. comparability studies are relatively younger and needs attention of people from different forward and easy to understand, so it was deemed appropriate for Pakistani audience where above which makes it a sensible choice for this research project. Because it is a straight application of Common Test Approach in the comparability studies are worthwhile as seen feasibility to a certain context (Newton et al., 2007). However, the results yielded from the have certain inevitable limitations and should be used with caution according to the comparability methods in use around the globe whether statistical or judgemental all GCE and A Levels every year (Tymms et al., 2005).

seen in more recent studies probing into issues like the comparability of the standards at Massey, 2000; Masssey et al., 2003, cited in Murphy, 2007). Common Test approach can be Local Examinations Syndicate (UCLES) undertaking comparability studies (Dexter & Newcastle and Durham. It has also been used by the researchers of University of Cambridge Curriculum, Evaluation and Management centre (CEM) located at the Universities of the comparability scene, for instance, it has been used until now by the researchers from the been declined in the 1970s but Murphy (2007) negates it and asserts that it was never out of different 20 examination boards in the UK was at its peak. Its popularity was thought to have notes, Common Test Approach saw its heyday when the demand for comparability studies of

procedures and content of Matric were evaluated, Matric English test was attempted to be aligned with CEFR after several meetings with high officials of FBISE.

Brannen (1992 cited in Cheng, 2005) while advocating the use of both the qualitative and quantitative methods asserts that these both methods are mutually dependant. By using both the methods makes the conclusion drawn from different data more confident and thus more reliable. Thus, cross-examination mechanisms provide a more confident conclusion drawn from different types of data which was desired for this study.

The two qualitative and quantitative methods also complement each other in this type of study. Phase 1 of the study compliments the latter stage of the study and substantiates the research assumption that the English testing systems of O Level and Matric are not equivalent. The quantitative strand of the study i.e. Phase 1 tests the hypotheses of the study with quantitative procedures by comparing the previous scores of the participants with the new scores gathered from the CTA (see section 4.5 above for Research Framework). Phase 2, on the other hand a qualitative approach, tests the research assumption i.e. the linking to CEFR gives the evidence of the content validity and the standing of Matric in an international framework.

Although, Chen (2005) referring to Bryman (1992), Cohen (1976), Cohen and Manion (1989), Halfpenny (1979), Hammersley and Atkinson (1983) and Patton (1987) suggests that qualitative and quantitative approaches both have their own strengths and weaknesses, the qualitative approach is usually time consuming, but results provide a more accurate picture of reality revealing more complexities than the results from the quantitative data. However, the results from this approach are often not generalizable. Quantitative approach, on the other hand, generally involving a large number of participants and a large volume of data provides results which can be generalized and are often over simplified but

show poor ecological validity. While the hypotheses of the study were tested by the quantitative methods in Phase 1, the inferences derived were from the qualitative methods in the Phase 2 of the study that the two testing systems; O Level's and Matric's English exams are not equivalent. So, the reason for combining both the approaches was because of the arguments like Brannen's (1992) which claims that the qualitative and qualitative research methodologies are both dependant and thus inseparable.

4.7 Sample of the Study

For the first phase of the study, the sample consisted of students from 5 private schools in Islamabad, Pakistan, who had studied at the same educational institution at secondary level, so that to have a homogeneous group in terms of subject background, having received similar teachers and teaching methodologies and having similar socioeconomic background (to the extent that this is reflected in their school choice). Thus, the controlled variables for this study were: teachers (the selected schools use the same teachers for the students of both the secondary education streams i.e. Matric and O Level), and the socioeconomic background of the participants. They were all the students of A Level/ 1st year FA/FSc⁴⁸ and had their official previous scores from the last exam they took (O Level/Matric). They sat in March 2012 for Matric and May-June for O Level. The researcher chose private schools because both the streams (Matric and O Level) are available only in the private schools apart from some of the very few semi government schools, so such a comparative study was only possible in the chosen private sector schools. The sample for the second phase of the study included Matric question papers of English from the years 2008-2012.

181 students, 83 Matriculates and 98 O Level graduates, were taken as a sample for the first phase. The ages of the participants of the study ranged between 16-18 years.

⁴⁸ FSc and A Level are considered equivalent levels of qualifications (IBCC, 2013).

Initially, an equal number of participants i.e. 100 from each gender were included but because of the incomplete data, only 100 males and 81 females were included in this study for the data analyses. The sample included students of variable academic achievement in English subject e.g. the Matric students' scores in English exam varied from 96% to 57% while O Level ranged from grades A* to D which were translated into numerical scores to have Matric equivalent scores according to the scale provided by IBCC and came to around 90% to 55%. Furthermore, the data was considered twice; first all the raw data from all the participants making the *unmatched group*. Then the participants were closely matched for all the variables especially their ages and secondary school achievement scores to make a *matched group*.

4.7.1 Unmatched group. For the comparison between two groups t-test is the appropriate statistical procedure (Bachman, 2004). So, in order to identify if the two groups differed on their age, gender, socioeconomic status (SES), t-test was used. Table 5 below demonstrates the age, gender, SES and the scores of the unmatched group. The data in Table 5 is called unmatched here because some of the data was discarded to completely match the groups in all the variables which yielded matched groups for the in depth analysis. The analysis was done for both matched and unmatched groups to have wider range of data analyses and to see if the analyses from both the groups yield the same results for a more valid conclusion. The matched groups included participants who were closely matched for their scores and age (see section 4.7.2 on matched group below for a detailed discussion on the matched group). The raw data included sample with the mean age of 16 years and 10 months for Matric and 17 years and 1 month for O Level. The participants included 44% females for both Matric and O Level. Their SES was also almost the same and finally their

attainment on their previous tests of Matric and O Level was also the same with a slight variation i.e their scores were 81.4% and 80.4% respectively.

	Matric	O Level
Age	16;10	17;01
Gender (% female)	44%	44%
SES	2.4	2.5
Matric Score*	81.4	80.7

Table 5: Unmatched group *Matric raw numerical scores were used while for O Level their grades were converted into numerical scores according to IBCC conversion scale.

4.7.2 Matched group. In order to realistically compare the two groups, it was essential to have both the groups equivalent in their gender, ages, schools, socioeconomic status (SES) and their secondary school achievement scores. They indeed were similar in their gender, schools (similar 5 English medium private schools with similar fee structure and curriculum), and socioeconomic status (according to their school choice and the SES questionnaire, see section 4.9) but there were differences in their ages and secondary school achievement scores. So to remove those differences some of the older participants and high achievers from O Level group were removed for the matched group t-test which resulted in a group not significantly different in age, gender, SES and their previous official test scores. Table 6 below demonstrates the age, gender, SES and their scores of the matched group used for the t-tests. They are identical in their average ages i.e. 16 years and 10 months. The participants included 44% females for Matric while 46% for O Level. Their SES was also almost the same with only .1% difference and finally their attainment on their previous tests of Matric and O Level was almost the same with a difference of only 1%, so their scores were 81.4%

and 80.4% respectively. The results from t-tests also suggested that the difference between the groups in terms of the above variables was not significant.

	Matric	O Level
Age	16;10	16;10
Gender (% female)	44%	46%
SES	2.4	2.5
Matric Score*	81.4	80.4

Table 6: Matched groups: *Matric raw scores were used while for O Level their grades

were converted into numerical scores according to IBCC conversion scale.

The decision to be made was to whether include all the data for the regression analysis or to include only the matched groups. The former approach was adopted for two reasons. Firstly, it produced a wider range of comparison. The larger number of counts gave greater precision and accounts for different levels of performance. Secondly, it enabled the full weight of data of students from different levels and took into account all the levels in relation to the students' performance on all the three tests and not just some specific levels thus gave a comparison on a wider range of levels.

However, for completeness, regression Analyses were also performed on the matched groups which were closely matched for their scores and ages, as also reported below.

4.8 Raters

The raters used for the second phase of the study were 5 certified examiners by world reputable organization and were the best choice to mark the subtest of IELTS chosen for this study. All the examiners who were willing to participate were included in the study. All the raters are currently working with different reputable international organizations. All have an

extensive experience (almost more than 20-25 years) in English language teaching at different levels and in different capacities. Rater 1 has English language teaching experience at school level but currently is only working as an examiner. Rater 2 also has taught at different levels currently teaching EFL at a private institute. Rater 3 on the other hand, taught at different levels especially 16+. She was also a key figure in devising the new English curriculum, Pakistan in 2006 for Secondary and Higher Secondary level. Rater 4 as well has an extensive background in English Language teaching at higher secondary and university level. Rater 5 has taught at different levels to different age groups starting from school to university. Only one of the examiners is a native speaker of English while the rest are non-native but have a native like competence evident from their being able to work as an examiner for international examinations which is a pre requisite for an examiner. The decision of including native speakers of English was only subject to the availability and most importantly the willingness of the examiners to be part of the study.

4.9 Research Tools

The test instruments of this study included the Manual CEFR (Council of Europe, 2009) and three English test batteries (Matric, O Level and the Common Test; reading and writing test modules of IELTS General module taken from the published material, 2002). Matric and O Level are described in detail with their scoring procedures in Chapter 3. Scores from these three test batteries were used in the first phase of this study. Samples of O Level and Matric English test papers are given in the Appendix B and D respectively.

The study attempted to measure the students' proficiency of English reading and writing level with the help of an English language proficiency test i.e. IELTS. As all the tests have limitations, for example, Hughes (1989) thinks the existence of "one best test" is a myth and therefore impossible to have. So, IELTS was thought to be the best choice to be selected

as a Common Test because it is a world recognized proficiency test (IELTS, 2013⁴⁹) and seemed most relevant to the purpose of the study i.e. to evaluate the language proficiency of the sample.

It is a challenging task to make comparisons between different tests as Taylor (2004) notes because different tests differ in their purpose, format and design. The testees might perform better in one than in others because of varying aptitudes but in this case both the groups of the sample had scores from two achievement tests (which are thought to be equal by IBCC (see Chapter 3, section 3.11.8) and the comparisons are made between the scores awarded by both the testing systems and how well they correspond to CEFR.

An objection that one might raise can be of the nature of different test batteries used in this study i.e. Matric and O Level are both achievement tests while IELTS, a proficiency test. Some might argue that since the tests are different in their types will yield different results and thus not comparable but as Council of Europe (2001:185) argues that the aim of both the test types i.e. “achievement (oriented to the content of the course)” and “proficiency (oriented to the continuum of real world ability)” is the same, that is to assess the linguistic abilities of a testee to be able to use in the real world. The difference between both the tests, as very aptly put by Kluitmann (2008), is that the achievement test looks backwards in respect of what the testee has achieved in a prescribed course while the proficiency test tends to look forward in regards to what the testee can do in the future and tells where he/she stands, making the results more transparent and comprehensible in terms of the testee’s linguistic abilities.

Another possible objection might be of the choice of the Common Test. Although, the Common Test should be appropriate to the level of the students of both the streams in terms of their age which in this case was 16-17 years. IELTS, on the other hand, as many think is

⁴⁹ See www.ielts.org

primarily designed for 18+ year olds mostly as a condition for university entrance or job opportunities but it was thought to be the best possible alternate of measurement because of many reasons. One, because of its familiarity and popularity with the student base in Pakistan, they know the value of IELTS and are happy to spare a couple of hours for some expert judgment on their performance on such a test which they consider to be useful in their future academic or professional endeavors. Two, because of its international recognition and the evidence of its validity and reliability available. Three, the availability and the consent of the Cambridge certified examiners to help with marking the scripts was an added and valuable advantage for the standardized marking of the Common Test. So, the assumption was that if IELTS is not very appropriate for 16+ year olds, then both the groups should do equally bad on it. Out of the two IELTS test formats available; Academic and General Training, only General Training format was used because the first writing task of both the versions differ. The first writing task of the Academic format is a graph/chart to be decoded by the candidates in order to write a report while the General Training format requires the candidates to write a letter. As the students of both the streams are quite familiar and trained in writing letters rather than writing reports based on tables and graphs, therefore, the General Training format was selected for this study. Another reason for the selection of the General Training format was that in its preparation literature about the format of the test (IELTS, 2002: 1), it states that General Training is meant for the candidates who wish to study at a non degree programme e.g. wishing to undertake a vocational or training programme, and admission to secondary schools thus not exactly a test meant only for 18+ year olds as generally considered.

A simple questionnaire adapted from the World Health Organisation (WHO) was also used to determine the socioeconomic background of the subjects (see Appendix F).

Although, studying at a similar private school was an indication of their similar economic background but to substantiate the fact even further this additional questionnaire was used. The questionnaire was also used to acquire information on subjects' age, gender and their Matric/ O Level official scores.

4.10 Decisions on Data Inclusion

The questionnaires and the reading and writing tests were sorted. Any incomplete test was not included in the analysis. All the raw data from all the participants was used for the analysis of unmatched group while for the matched group it was essential to have both the groups equivalent in their gender, ages, schools, socioeconomic status and their secondary school achievement scores to realistically compare the two groups and thus some of the data had to be removed to get the required results.

4.11 Procedure of the Study

As mentioned earlier, the study was divided into two phases; phase 1, dealt with the quantitative strand of the study where the Common Test Approach was adopted in order to compare the scores of students from Matric and Level in English. While, in phase 2, an attempt was made to align Matric English test to CEFR.

4.11.1 Phase 1

In the first phase, for the reliability measure of their previous O Level / Matric test scores and the IBCC conversion formula, the students of A Level, in the first year of their A Level/ 1st year FA/FSc took a Common English Test (reading and writing modules of IELTS) to assess the students' proficiency in reading and writing and to see how meaningfully the scores can be compared. Their scores from O Level/Matric could be compared through the scores from the Common Test. The results from both the groups were compared within (previous scores compared with their scores on the Common Test) and

between the groups (O Level/Matric) to see if their new scores differed from their previous one and if the groups (O Level/Matric) differed from each other significantly.

Thus, the comparison of the scores was multi tiered where the difference of scores within and between the groups was considered, an important technique for in depth analysis. Firstly, the scores of both the groups from this test were compared between both the groups to see which group scored better. Secondly, the scores were compared with in the groups i.e. the scores of both the groups from the Common Test (subtest of IELTS) were compared with their respective previous English test scores in Matric and O Level and then the performance of both the groups was also compared.

The Quantitative Data was analysed with using t-test and Linear Regression illustrated in Chapter 5 in the form of tables and graphs for the graphic representation of group differences. Both the groups were compared using statistical method of t-test as suggested by Bachman (2004), the best possible statistical procedure for the comparison of two groups with independent scores. In order to establish the relationships between the scores of all three tests i.e. Matric; O-Level and the Common Test, linear regression was used, a statistical technique for measuring the relationship between the variables (Sykes, 1992; Bachman, 2004). Linear Regression helped in determining the differences within the groups as well i.e. whether the high achievers also performed higher on the Common Test.

Before taking the Common Test, the participants were asked to fill out the questionnaire to ensure the students similarity of socioeconomic background (see Appendix F).

The test was conducted in exam like conditions as is prescribed by IELTS in the presence of an invigilator where the participants were not allowed to talk or cheat from each

other. One hour was given for the reading test and one for the writing test. The researcher herself was present at all times of the data collection.

The reading test was marked by the researcher herself as it is more of a clerical and objective marking (the answer is either correct or incorrect) but the writing scripts which are more subjective in nature were distributed amongst five certified examiners (see section 4.8). The writing scripts were coded (participants' and school names were removed) so the examiners would not be biased against any group. They marked unaware of whether the scripts were from the Matric or O Level stream.

4.11.2 Phase 2

In the second phase for the validity measure, the whole Matric English testing system was analysed with the help of procedures provided by the Manual for relating language examinations to CEFR (Council of Europe, 2009). To see if the Matric English test could be aligned with an international framework like CEFR, the desirable situation would have been to involve a panel of experts i.e. teachers, paper setters and examiners/markers and go through the whole linking process as proposed by the Manual. But unfortunately, a whole panel of experts could not be involved; one because of the lack of resources and two it seemed logical to carry out a small scale study to see if it is feasible at all to align Matric test to CEFR, so. It was taken as a preliminary study investigating for the feasibility of a full fledged alignment process to be carried out. As the Manual repeatedly cautions the users to be sure of the quality of the exam or test to be linked to the CEFR because a qualitatively poor examination, the Manual warns is a wasted enterprise and would hence bear no fruit. The Manual (2009) suggests the pre-requisites of the test to be linked to the CEFR, the foremost being the ability of the test to demonstrate validity in its own right. The quality of testing procedures are skeptical if the test has no procedures for the examiners marking

reliability or the successive forms of the same test that are equivalent cannot be linked to CEFR. Then, it might be claimed that the test lacks the ability to exhibit its “internal consistency in the operationalisation of its standard(s)”. Therefore, through the interviews with the members of the FBISE, information was gathered whether there are procedures for determining the validity of the Matric English test. Upon investigation, it was revealed that there are a few measures carried out informally to ensure validity but any concrete report could not be shared. Hence, it was deemed appropriate to have a preliminary study of linking Matric to CEFR as far as possible with a couple of experts and if the results are not disappointing the relevant authorities in the government and the educational boards can take it more seriously and execute it on a large scale with a complete panel of experts and all the detailed procedures proposed in the Manual (2009).

In the circumstances of the non availability of a whole panel of experts, it was thought to be the best possible solution to have a preliminary study of linking Matric to CEFR with 2 certified examiners by reputable international organization to go through the whole linking process as proposed by the Manual. 2 skilled examiners carried out the alignment process. One was the researcher herself, a certified examiner with extensive English teaching and testing experience and the other examiner with extensive experience and training in examining international English tests recognized worldwide was requested to have the best expert opinion. She had also been involved in the English curriculum planning of Matric (2006). So, she was not only familiar with the CEFR and its descriptors through her affiliation with the international English examinations but also with the Matric English curriculum in depth.

The linking process to CEFR ideally includes workshops and detailed discussions formally classified by the Manual as familiarisation, specification and standardization. The

The second activity included self assessment in one of the foreign language known to them on the basis of CEFR Table 2 (CEFR, 2001:26, Portfolio self-assessment grid) and mark key words which described their ability best and also give a rationale as to why that level is different than its adjacent levels so to understand the salient features of the descriptors. Both the participants did the same activity and discussed in detail during their face to face interaction.

The third activity was the descriptor sorting task to classify all the jumbled up CEFR descriptors to their appropriate level and reconstruct CEFR Table A1 (see Appendix G). This activity was also used to reconstruct Table C4 (see Appendix H) of the Manual “Written Assessment Criteria Grid”.

Upon meeting, all these activities were discussed by the participants. Self assessment of activity two was discussed and the selection of one level and not the other was contemplated and justified. The answer key to the third activity was discussed and the salient features of all the levels were studied and agreed upon. The features that distinguish one level from the other were thoroughly considered and the minimum requirements for each CEFR levels were outlined. By the end of this stage a considerable and desirable familiarity to CEFR and the descriptors was achieved.

4.11.2.1.2 Specification. To provide the internal and external validity of the test in question this is a self audit stage of the alignment process which investigates the coverage of the examination content and task types. This stage required different forms to be filled i.e. Forms A1 to A23. These forms were filled with the help of all information available from the interviews with the F BISE staff, the Matric curriculum, question papers, and CEFR. This was certainly an awareness-raising activity about the quality of Matric examinations. During the process features of language use and competence covered by Matric were mapped, as

suggested by the Manual, to CEFR as in CEFR chapter 4 “Language use and language learner” and CEFR chapter 5 “The user/learner’s competences”. Forms A2 and A8-A20 focused on the content analysis of Matric and its consequent relationship to the CEFR. Forms in Appendix A and C of the Manual (2009) were also filled with mutual agreement of the examiners. Of course, there are quantitative measures of content analysis (e.g. Kaftandjieva, 2007) but this study followed the Manual and only considered the qualitative measures.

4.11.2.1.3 Standardisation. This stage makes sure that there is a common understanding of the CEFR levels with the help of illustrative samples of spoken and written performances available at www.coe.int/lang. Although, Matric does not have any speaking test but for the better understanding of the CEFR levels and its distinguishing features, the spoken performances were also exploited. Therefore, this stage deepened the familiarity with the CEFR levels. Form B3 of the Manual was filled in with the justification and rationale of the participant’s decisions using the key terminology from the CEFR scales.

Since, the participant were also Cambridge examiners for English, the bands of IELTS and their equivalent CEFR levels provided by ESOL, Cambridge (2011: 26) were also used to elucidate the CEFR levels even further by relating to the existing knowledge endorsed by Cambridge professional training.

Training in rating reading and writing performance using CEFR scales. Again, there was no formal training session as suggested by the Manual but the practice was carried out, as proposed in the Manual in a rather informal fashion e.g. *controlled practice* of scoring before the *individual assessment* of scripts by both the participants. CEFR scales were revisited and their corresponding sample writing performances were studied carefully and exploited to make sure that there was a clear and common understanding of the CEFR levels. The CEFR levels awarded to the already standardized sample performances were discussed

thoroughly. All the levels were closely studied focusing especially on their distinguishing features. The writing performances selected for the training purpose were taken from the Council of Europe's containing CEFR calibrated pilot samples⁵⁰ which used reading and writing tasks. The reading and writing tasks included post card, story, letter, report. The tasks and the sample performances ranged from A2 to C2 levels of CEFR which can be seen in Table 7.

CEFR level	Task	Response Type
A2	Daily life	Post card reading, writing personal letter.
B1	Daily life	Story writing
B2	Free time, entertainment	Letter reading, Letter writing using the notes.
C1	Daily life; relations with other people	Letter reading, report writing using instructions.
C2	Cultural affairs	Reading newspaper article, writing letter.

Table 7: Material used in the training rating activity

The participants read the samples and scored them individually using the appropriate scales e.g. Table C4 of the Manual "Written Assessment Criteria Grid" for writing (see Appendix H), which they had also used in the sorting task in the familiarization stage and for reading Tables A1, A2 and A3 (see Appendix G) for receptive/ linguistic texts and items were used. The scores were then shared and discussed as to why a specific level was awarded

⁵⁰ Samples available at http://www.coe.int/t/dg4/education/elp/elp-reg/Source/Key_reference/exampleswriting_EN.pdf accessed on December 20. 2012.

before finally consulting the right answer allocated to the already standardized illustrative scripts of written performances.

Benchmarking- rating of Matric sample English answer papers. As generally, the end product of training is the benchmarking so, after a mutual and clear understanding of the CEFR levels, local samples were related to the CEFR levels and scored by both the participants accordingly. The sample of English writing performances was requested from the Chairperson FBISE to which she agreed. The sample requested to range from high (i.e. the best paper) to low (very weak). They were supposed to be representing typical six different levels to correspond to the six levels of CEFR (85% and above, 75-84%, 65-74%, 55-64%, 45-54%, 44% and below). For each level, 4 illustrative samples were requested; 2 for Matric SSC I and 2 for SSC II. The writing samples were supposed to include papers from the annual board examinations of year 2012. Scores from the same year were also used for the first phase of this study.

As repeatedly assured by the Chairperson FBISE for the provision of Matric answer papers in the interviews initiated in December 2012 but by the end of consistent inquiry of 4 months, it was told by FBISE that giving away any part of the papers is a breach of their confidentiality policy and therefore would not be provided under any circumstances. Keeping in view, the non availability of the Matric sample answer papers and the determination not to terminate the process of alignment to CEFR at this stage, an alternative approach was adopted. Some schools were requested to provide samples from their pre-board examinations which are based on the Matric Annual Examinations' paper pattern, regularly administered in all schools before the final board exams. After, approaching a few schools, one of the schools was able to provide 24 English reading and writing answer papers; 4 for each 6 levels of CEFR (2 for Matric SSC I and 2 for SSC II). Fortunately, the teachers who marked them are

trained by FBISE and are regular examiners for FBISE Matric English papers. So, the alternative approach of using pre-board samples was the best substitute serving the purpose and goals of the research study.

The samples were marked in the same way as was done in the training stage. Both the examiners scored one script for each level together. After sharing the scores, individual rating followed using the appropriate scales e.g. Table C4 of the Manual “Written Assessment Criteria Grid” for writing, which they had also used in the sorting task in the familiarization stage and for reading Tables A1 for writing, A2 and A3 for receptive/linguistic texts and items were used.

Standard setting- rating of Matric reading and writing tasks. Thinking at the task and item level, both the participants tried to evaluate the level in terms of the difficulty of the task by answering the following question: “At what CEFR level can a test taker already answer the following task correctly?” The participants tried to think of the lowest possible competence level required to answer the task keeping CEFR levels in mind. Subsequently, the task rating form B5 of the Manual (2009) was filled in.

Materials used for this stage were the question papers of last five years i.e. from 2008-2012 for Matric English exam were used. The scales used for the standard setting were the same as used in the training e.g. Table C4 for writing and for reading Tables A1, A2 and A3.

CHAPTER 5

DATA ANALYSIS

This chapter discusses the results and analysis of the data. It provides the statistical analysis of the data for the first phase of the study and finally the results derived. Latter part of the chapter provides data analysis for the second phase of the study.

5.1 Phase 1

5.1.1 Data Analysis

As can be seen in Table 8 below, the O Level group out performs the Matric group on the Common Test. Matric students' performance ranges from 5.9 to 6.3 on the scale of 9 (of IELTS sub test) for the sub criteria for reading and writing while O Level students performance ranges from 6.9 to 7.3 on the same sub criteria. The average performance also differs for both the groups as Matric students average performance on the Common Test is 6.1 whereas O Level out performs them by one point which is significant for the scale of 9 points. In fact, even half a point can restrict a candidate's academic and professional opportunities. According to the t-tests as can be seen in Table 8 and its graphical representation in Figure 3 and 4, the O Level group clearly outperformed the Matric group on the Common Test which is further substantiated by the linear regression analyses below. The

p-values for all the criteria of the Common Test showed highly significant differences (with $p < 0.001$).

	Matric	Olevels
CRT	6.1	7.1
CWT_1.1	5.9	6.9
CWT_1.2	6.1	7.1
CWT_1.3	6.2	7.2
CWT_1.4	6.0	7.1
CWT_2.1	6.0	6.9
CWT_2.2	6.0	7.1
CWT_2.3	6.3	7.3
CWT_2.4	6.0	7.2
CWT_1_Average	6.1	7.1
CWT_2_Average	6.1	7.1
CWT_Average	6.1	7.1
CR&W_Average	6.1	7.1

Table 8: Scores of Matric and O Level on the Common Test.

Key: Common reading test (CRT), Common writing-task-1-criterion-1(CWT_1.1), task-1-criterion-2 (CWT_1.2), task-1-criterion-3 (CWT_1.3), task-1-criterion-4 (CWT_1.4), Common writing-task-2-criterion-1(CWT_2.1), task-2-criterion-2 (CWT_2.2), task-2-criterion-3 (CWT_2.3), task-2-criterion-4 (CWT_2.4), Average writing-task-1 (CWT_1_Average), Average writing-task-2 (CWT_2_Average), Average writing score (CWT_Average), Average reading & writing score (CR&W_Average).

The same information of attainment of the two groups on the Common Test can be seen graphically below in Figure 4, where the O Level students outperform the Matric students consistently. There is a constant high performance of almost 1 point across all the criteria for reading and writing on the scale of 9 of the subtest of IELTS.

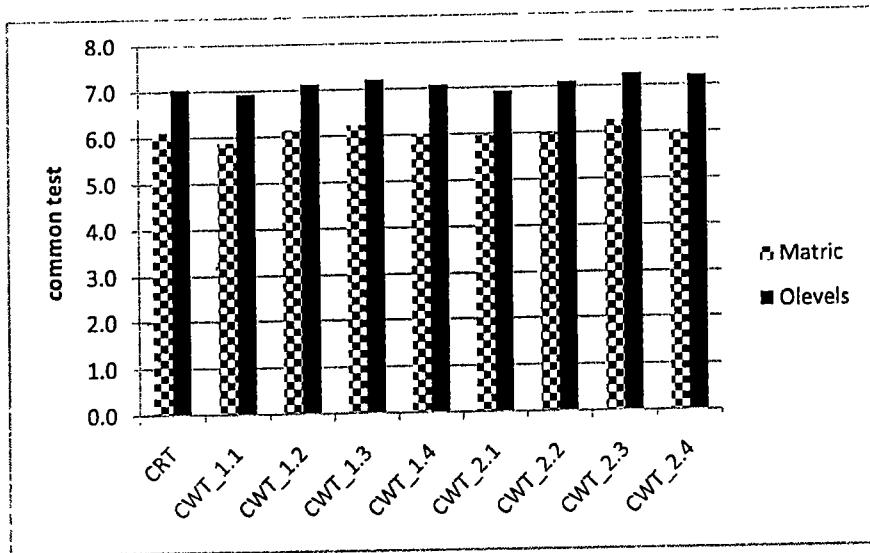


Figure 4: Scores of Matric and O Level on the Common Test.

Key: Common reading test (CRT), Common writing-task-1-criterion-1(CWT_1.1), task-1-criterion-2 (CWT_1.2), task-1-criterion-3 (CWT_1.3), task-1-criterion-4 (CWT_1.4), Common writing-task-2-criterion-1(CWT_2.1), task-2-criterion-2 (CWT_2.2), task-2-criterion-3 (CWT_2.3), task-2-criterion-4 (CWT_2.4).

Again a very positive and constant trend of high attainment on the Common Test of O Level students if only the averages across all the criteria are considered as can be seen in Figure 5 below.

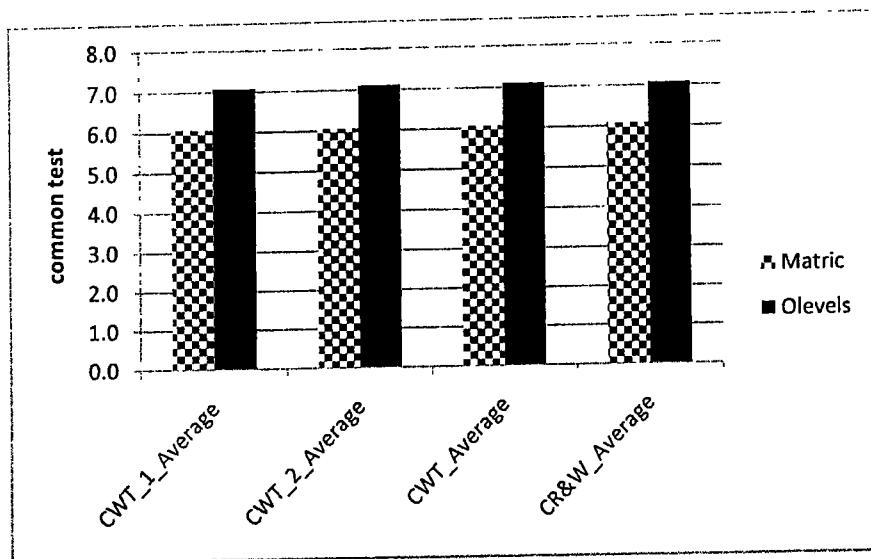


Figure 5: Average Scores of Matric and O Level on the Common Test.

Key: Average writing-task-1 (CWT_1_Average), Average writing-task-2 (CWT_2_Average), Average writing score (CWT_Average), Average reading & writing score (CR&W_Average).

5.1.2 Linear Regression analysis

Regression analysis is first done for the unmatched group where all the data is used and then the same is done for the matched group to see if that made any difference when the sample is matched for age and scores. The regression lines model the participants' performance on the new test as a function of their performance on the test they sat i.e. Matric or O Levels. Using the scores from the Common Test we can bridge between the scores from O Levels to Matric.

For the first Regression analysis, raw Matric scores and the numerical version of the O Level scores are used for the unmatched group. Through the regression analysis, a conversion scale is yielded for both O Level and Matric i.e. by using their scores from the Common Test and their respective Matric and O Level. Regression analysis suggests that the

optimal conversion from Matric scores to scores on the Common Test is achieved by the formula $C = 0.0545M + 1.66$, where C is the score on Common Test and M is the score on Matric. The optimal conversion from scores on O Level, normalised to the Matric scale (A* = 90+, A=85+, B=75+, C=65+, D=55+,E=45+), to scores on Common Test is given by the formula $C = 0.0516L + 2.86$, where C is the score on Common Test and L is the score on O Level.

Treating the O Level scores as if they were Matric scores (as per the current one-to-one conversion procedure by IBCC), it can be examined whether this accurately predicts participants' performance on Common Test. If we apply the Matric-to-Common Test conversion formula to these scores, we obtain the following estimates of performance on Common Test: 55 (D) corresponds to 4.66, 65 (C) to 5.20, 75 (B) to 5.75, 85 (A) to 6.29, and 90 (A*) to 6.56. Comparing these predictions to the participants' actual performance on Common Test, it is found that this underestimates their performance by an average of 0.97 points, roughly equivalent to just under two grades.

Out of the 95 O Level students considered in this experiment, the present conversion underestimates the performance on Common Test for 90 students, and overestimates performance for just 5 (a highly significant difference; $p < 0.001$, sign test). The greatest overestimation for a single student is of 0.62 points; the greatest underestimation is 1.96 points.

If the Common Test is considered to be an accurate assessment of students' competence in English, it is clear that the current conversion system undervalues the performance of O Level students by comparison with Matric students. It is, however, hoped that the results of this experiment would suggest a fair conversion system (derived from the regression analysis below) of $M = 0.947L + 22.1$. This indicates that the competence of

students gaining D-grades in O Level is (on these data) roughly equivalent to that of students achieving 74.1 on Matric, that a C-grade corresponds on average to a score of 83.7, a B to 93.1, and A and A* lie above the range of the Matric scale.

However, these results should be interpreted with caution: the linear relation that is posited is likely to break down at the extremes of the range, and further study would be needed to ascertain precisely how the O Level grades should be interpreted. The current findings suggest that the current ratings of O Level grades in this subject are not appropriate. It can be stated with high confidence that their valuations in comparison with the Matric scale should be increased by at least 5 points, pending further investigation.

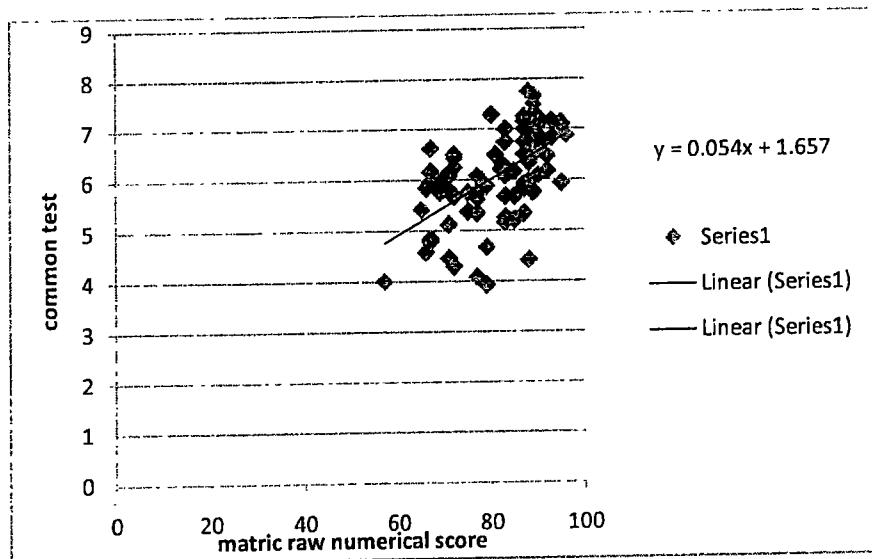


Figure 6: Linear regression analysis for Matric versus Average score (Unmatched group)

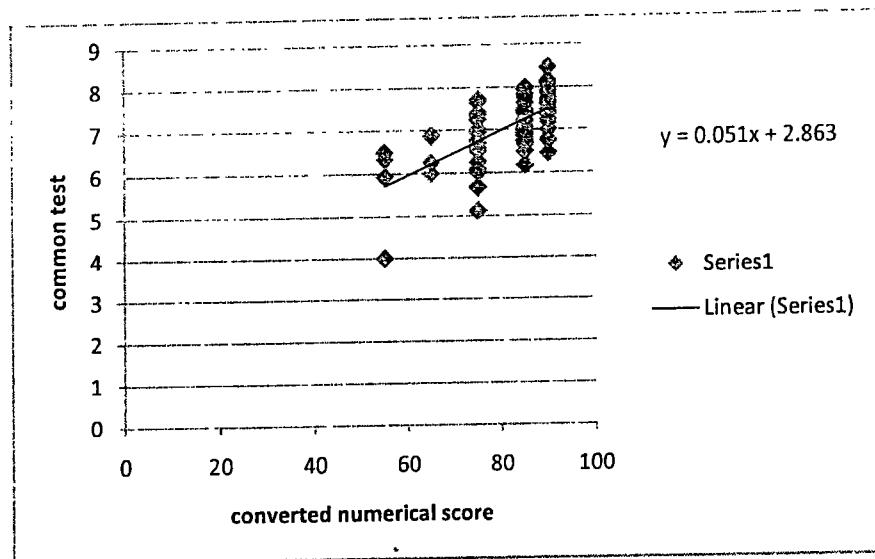


Figure 7: Linear regression analysis for O Level versus Average score (Unmatched group)

Regression analysis for the matched groups yielded the same results as for the unmatched and can be seen below in Figures 8 and 9. The conversion factors calculated through regression analysis seems the best fit for this study. The expectation can be derived that a student who gets 6 on the Common Test would get 76.1 on Matric and 60.8 on O Level, and one who gets 7 on the new test would get 96.3 on Matric and 80.2 on O Level, and so on. It can be concluded with greater confidence about the accuracy of this conversion, on average, towards the middle of the studied range, as the data here is more extensive. Also, this analysis presupposed a linear relation between the scores on the tests, which may break down at the extreme values due to ceiling effects.

5.1.3 Analysis for matched groups

For the second Regression analysis, again un-rounded raw numerical scores of Matric and the numerical version of the O Level scores are used for the matched groups. Regression analysis suggests that the optimal conversion from Matric scores to scores on Common Test is achieved by the formula $C = 0.0545M + 1.66$, where C is the score on the Common

Test and M is the score on Matric. The optimal conversion from scores on O Level, normalised to the Matric scale ($A^* = 90, \dots$), to scores on Common Test is given by the formula $C = 0.0536L + 2.70$, where C is the score on Common Test and L is the score on O Level.

Treating the O Level scores as if they were Matric scores (as per the current one-to-one conversion procedure of IBCC), it can be examined whether this accurately predicts participants' performance on the Common Test. If Matric-to-Common Test conversion formula is applied to these scores, the following estimates of performance on Common Test are obtained: 55 (D) corresponds to 4.64, 65 (C) to 5.19, 75 (B) to 5.73, 85 (A) to 6.28, and 90 (A^*) to 6.55. The conversions are almost the same as the results from the unmatched group.

Comparing these predictions to the participants' actual performance on the Common Test, it is concluded that the current method underestimates performance on the Common Test by a mean of 0.97 as was found in the analysis for unmatched group. For the 80 students considered for the matched group analysis, it is found that the present conversion formula underestimates the performance on Common Test for 75 of them, and overstates 5 of them (a significant difference; sign test, $p < 0.001$). This closely conforms with the earlier results from regression analysis of the unmatched group.

Again a similar conclusion can be drawn for this matched group as for the unmatched above that if the Common Test is considered to be an accurate assessment of students' competence in English, it is evident that the current conversion system of IBCC undervalues the performance of O Level students by comparison with Matric students. The results of this experiment would suggest a fair conversion system (derived from the above regression analysis of the matched groups) of $M = 0.983L + 19.0$. This indicates that the competence of

students gaining D-grades in O Level is (on these data) roughly equivalent to that of students achieving 73.0 on Matric, that a C-grade corresponds on average to a score of 82.9, a B to 92.8, and A and A* lie above the range of the Matric scale.

Thus, there is not much of a difference between the regression analysis of the matched and the unmatched groups. It is concluded that the current method underestimates performance on the Common Test by a mean of 0.97 roughly equivalent to just under two grades of O Level which is quite substantial. On the basis of the current findings, it can be stated with high confidence that the current ratings of O Level grades in this subject are not appropriate, and would suggest that their valuations vis-a-vis the Matric scale should be increased by at least 5 points, pending further investigation.

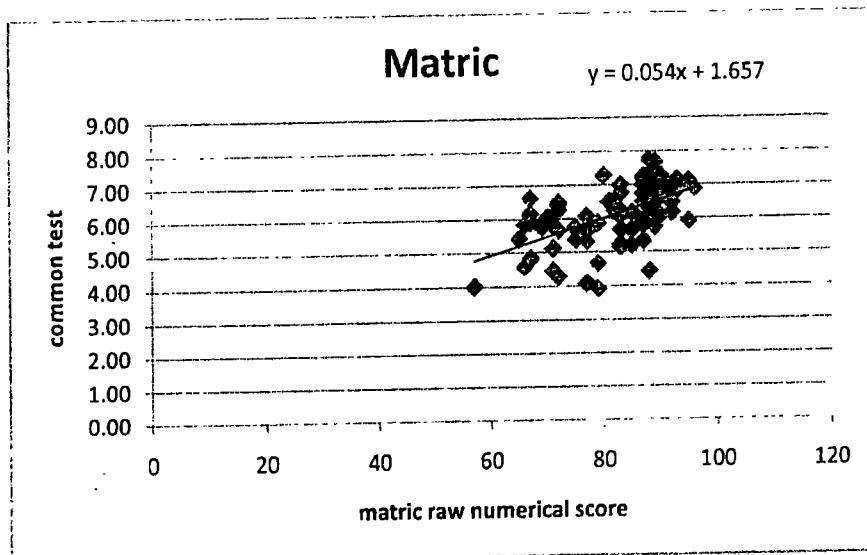


Figure 8: Linear regression analysis for Matric versus Average score (matched group)

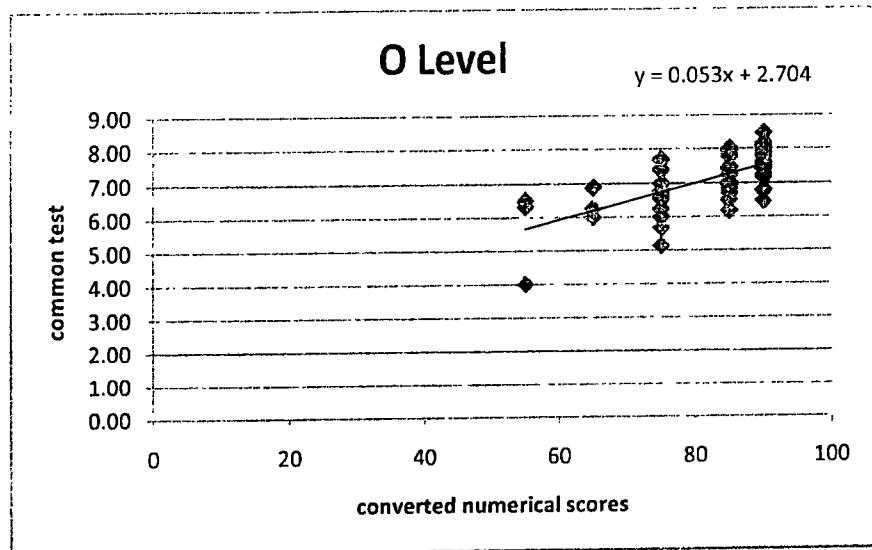


Figure 9: Linear regression analysis for O Level versus Average score (matched group)

The current conversion (implicitly 1 to 1) can be tested by applying it to the O Level group, that is, their scores to be considered Matric scores, and can be seen whether their performance on the Common Test is in line with these Matric scores. So for each individual their scores are estimated on Common Test based on the Matric-to-Common Test conversion factor, and compared to their real score on Common Test. The differences (residuals) are plotted in the error graph below (Figure 10). The average difference is about 1.0 (that is, the real Common Test score is about 1.0 higher than would be expected, on average), and almost always greater than zero (that is, the real Common Test score is higher than would be expected for the vast majority of participants).

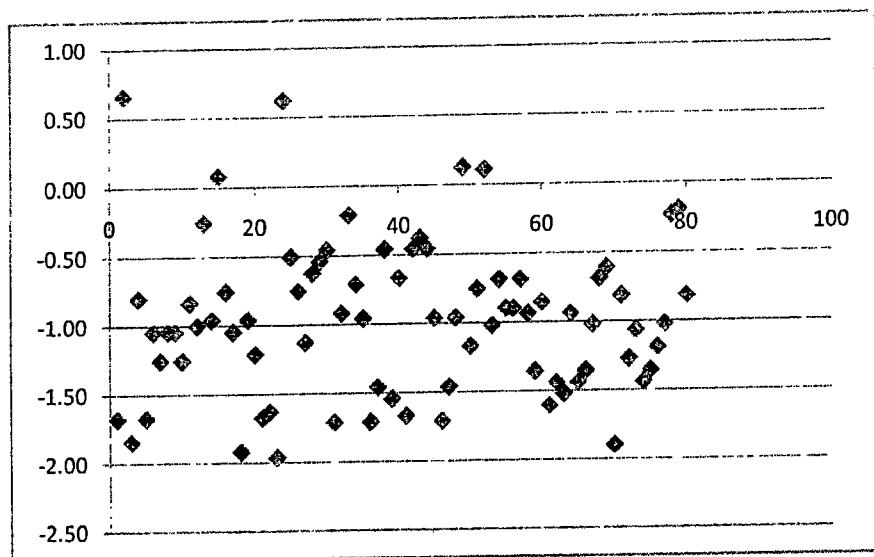


Figure 10: Distribution of errors on current conversion

It can also be considered how the O Level students tested were individually expected to perform on the Common Test, according to the current conversion system and according to the proposed conversion system. Plots of the errors can be seen above as Figure 10 and Figure 11 below for the two conversion methods. Figure 10 shows how the majority of O Level students are predicted worse performance on the Common Test, under the current conversion system, than they in fact exhibited. The new method, by definition, estimates mean performance correctly on the Common Test; the second plot shows the distribution of the resulting errors around zero.

Based on the regression lines, “fair” conversion rate is calculated – that is, one that would make unbiased estimates of the performance on the Common Test. The error graph as seen in Figure 8 shows how that would work: it estimates the score for each individual based on a two-step process of (a) converting their score to Matric by the proposed conversion rate and (b) converting this score to Common Test. (This is actually just the same as using the direct O Level-to-Common Test conversion rate, mathematically). Here the mean of the

errors is zero, although naturally this overstates some participants' performance at the same time as understating that of others, because O Level and Common Test results are not perfectly correlated. (In fact, these errors are just the same as the deviations from the line of best fit in the O Level versus Common Test graph, see Figure 9)

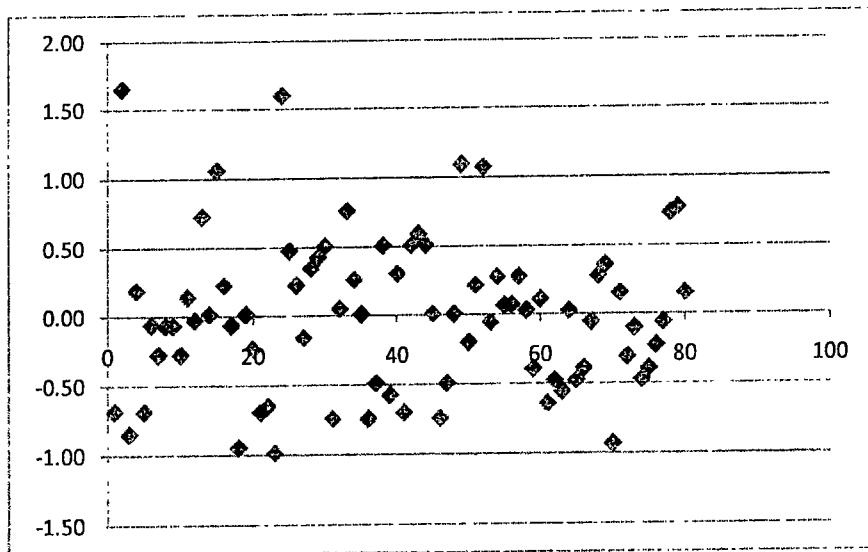


Figure:11 Distribution of errors on proposed conversion

Table 9 below presents the discrepancy between the converted scores of O Level by IBCC and the proposed formula. Figure 12 below represents the discrepancy graphically. As can be seen, according to the evidence provided, it is proposed to revise the IBCC formula at least for the English which is the focus of this research study. This indicates that the competence of students gaining E-grades in O Level is (on the basis of the empirical data) roughly equivalent to that of students achieving 63.3 on Matric, that a D-grade corresponds on average to a score of 73.0, a C to 82.9, a B to 92.8, and A and A* lie above the range of the Matric scale. Thus, on the basis of students' performance on the Common Test which is a subset of IELTS (considered a reasonably accurate assessment of students' competence in English) it is concluded that the current conversion system undervalues the performance of O Level students by comparison with Matric students.

Grade	converted marks proposed formula	converted marks IBCC
A*		90
A		85
B	92.8	75
C	82.9	65
D	73	55
E	63.3	45

Table 9: O Level grades converted according to the proposed and IBCC formula

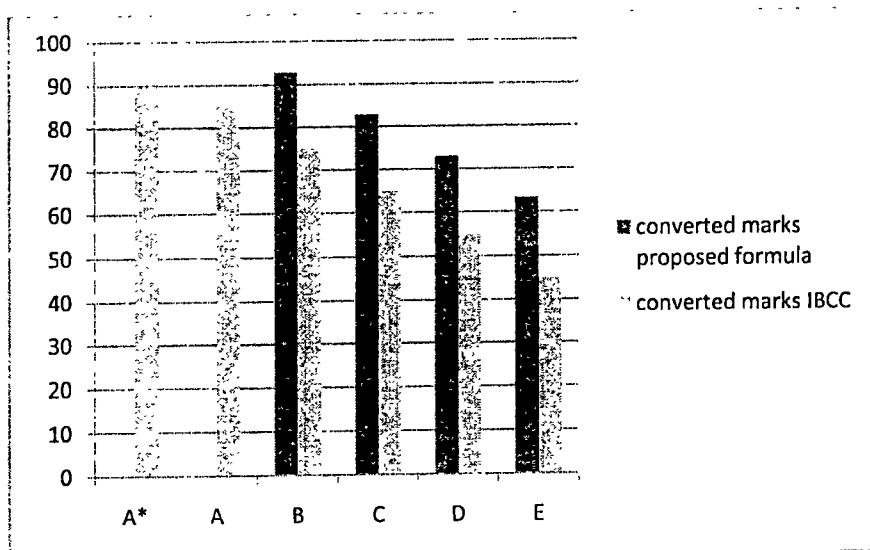


Figure 12: O Level grades converted according to the proposed and IBCC formula

5.1.4. CEFR and Matric Scores

Before attempting to align the content of Matric English test to CEFR, in the phase 1 with the help of the data from Matric English scores and the scores from the Common Test in

in this study an attempt was made to see the equivalence of Matric scores on the CEFR six levels. The relationship between IELTS and CEFR cannot be one-to-one as IELTS employs a 9 point band scoring system while CEFR has 6 levels as can be seen in Figure 13 below.

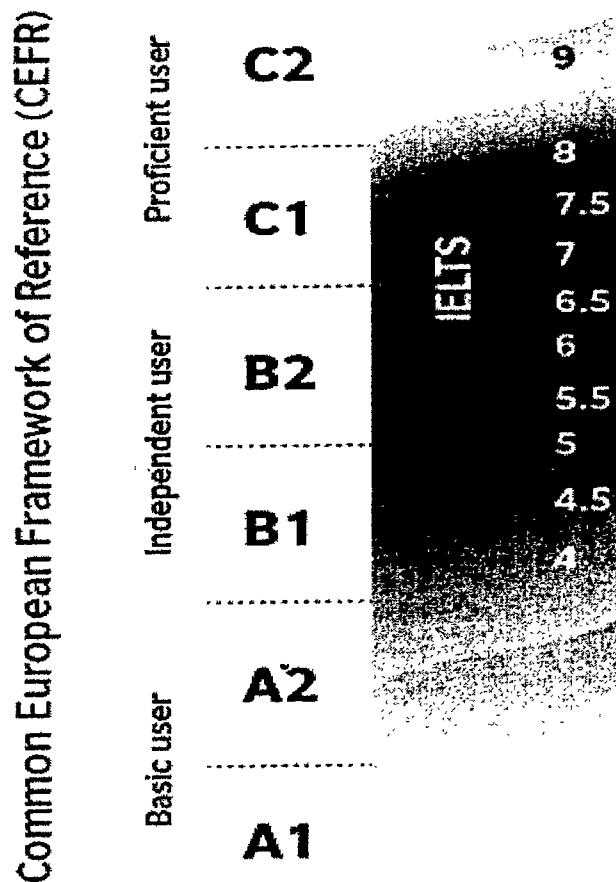


Figure 13: the mapping of the IELTS scale to the CEFR above derived from the interrelationship between IELTS and the Cambridge English Language Assessment Main Suite qualifications and the mapping of these latter qualifications to the CEFR. For further information see http://www.ielts.org/researchers/common_european_framework.aspx

Using the IELTS to CEFR equivalence table (IELTS being the Common Test in this study) we can compute to have an estimate of someone's performance on IELTS as a

function of their Matric performance ($C = 0.0545M + 1.66$). So, we can plug in a Matric score M , read off the expected IELTS score C , and match that against the scale. The grade boundaries would be midway between the numbers given, in principle, so 5.25 or thereabouts for B1/B2, 6.75 for B2/C1, 8.25 (or 8.5 approx.) for C1/C2.

Based on this model, one would need to score 65.9 on Matric to reach B2, 93.4 to reach C1, and the scale does not extend as far as C2 as can be seen in Table 10 below.

	CEFR	IELTS	Matric
Proficient User	C2	8.25-8.5	93.4
	C1		
Independent User	B2	6.75	65.9
	B1		
Basic User	A2	5.25	
	A1		

Table 10: Relationship of CEFR levels, IELTS and Matric scores

5.1.5 Discussion: The Conversion Scale

Applying the Common Test approach, this research shows that the current conversion scale used by the IBCC undervalues the performance on O Level students of English by comparison with Matric students. Based on the regression analysis above, the results suggest a fair conversion system (derived from the above regression analysis of the *matched group*) of $M = 0.983L + 19.0$. This indicates that the competence of students gaining D-grades in O

Level is roughly equivalent to that of students achieving 73.0 on Matric, that a C-grade corresponds on average to a score of 82.9, a B to 92.8, and A and A* lie above the range of the Matric scale.

While, the results of the regression analysis from the *unmatched group* suggests a fair conversion system of $M = 0.947L + 22.1$. This indicates that the competence of students gaining D-grades in O Level is roughly equivalent to that of students achieving 74.1 on Matric, that a C-grade corresponds on average to a score of 83.7, a B to 93.1, and A and A* lie above the range of the Matric scale.

As evident, there is not much of a difference between the regression analysis of the matched and the unmatched groups. It is, therefore, concluded that the current method underestimates performance on the Common Test by a mean of 0.97 roughly equivalent to just under two grades of O Level which is quite substantial. The current findings strongly suggest that the current ratings of O Level grades in this subject are not appropriate. We can state with confidence that their valuation on the Matric scale should be substantially increased (probably by 5 points), but the precise extent to which this should be done will require further investigation to determine.

5.2 Phase 2: Results of Alignment of Matric English Exam to CEFR

The data from Phase 2 was analysed with appropriate qualitative and quantitative procedures as discussed below.

5.2.1 Familiarisation

The third activity at the familiarisation stage was the descriptor sorting task to classify all the jumbled up CEFR descriptors to their appropriate level and reconstruct CEFR

Table A1 and Table C4. The participants were successful in placing the descriptors in their appropriate CEFR level.

Although, there were 2 instances where the placement was not correct but was discussed until the point of agreement was achieved.

5.2.2 Standardisation

5.2.2.1 Training in rating reading and writing performance using CEFR scales.

Following Papp and Salamaura (2009), the labels representing the CEFR levels were converted into a numerical scale as presented in Table 11 in order to statistically analyse the responses of the examiners.

CEFR level	Scale
A1	1
A2	2
B1	3
B2	4
C1	5
C2	6

Table 11: Conversion of labels representing CEFR levels into a numerical scale

5.2.2.2 Benchmarking- rating of Matric English answer papers. The examiners after trained in the use of CEFR levels and its descriptors, individually scored a script at each 6 levels. This step of the procedure was carried out in the same way as the training i.e. the scoring followed the sharing of the scores and then ended with a discussion. During benchmarking there was high rater agreement for the scripts used. The examiners generally agreed on the similar score to be awarded to the writing scripts.

5.2.2.3 Standard setting- rating of Matric reading and writing tasks. The most frequent ratings were A1/A2 of CEFR levels for the reading tasks. The reading comprehension tasks mainly required the candidates to reproduce which had been learnt prior to the exams from the text books. As discussed in chapter 3 (see section 3.11.5.4 on Matric English papers), the reading tasks had no reading passage given to consult for answering the questions, thus requiring the candidates to use their memory and reproduce the text read in the classroom. On average, the Matric English reading tasks rating for the papers of last 5 years 2008-2012 did not go beyond A2 while for writing tasks, B1-B2 and occasionally C1 was agreed upon for the last section of the writing tasks e.g. letter writing or essay writing.

5.2.2.4 Standard setting- rating of Matric reading and writing sample answer papers. The ratings of the examiners generally focused only on the writing component of the Matric writing and reading sample answer papers as the reading component which mainly required reproduction of memorized responses and did not test Reading Skills of the test takers. The joint decision of the examiners was that the reading skills that are required and elicited in Matric Reading section of English paper could not be judged as there is no evidence of dependence on memory in the CEFR scales. Such like question tasks might be considered as pre-A1 level and therefore could not be judged with the present 6 levels of CEFR. However, the writing sections were rated and scored according to the CEFR 6 Levels (A1-C2).

The scores from the *individual assessment* by both the participants were statistically analysed. They scored using the CEFR 6 level scale, which was then converted according to the numerical scale (from 1-6) in Table 11.

Matric scores of the English writing sample answer papers rated originally by the teachers were divided into six different levels to correspond to the six levels of CEFR (85% and above, 75-84%, 65-74%, 55-64%, 45-54%, 44% and below) and to be statistically

analysed. The six divisions were converted into a numerical scale as presented in Table 11 the labels representing the CEFR levels in order to compare with the responses of the examiners and the original score awarded by the teachers.

Figures 14 and 15 below graphically presents the scores awarded by the Cambridge certified examiners and the teachers for Matric SSC I and SSC II respectively. There is a mutual agreement between the examiners on most of the scripts. None of the scripts scored by the examiners could achieve more than level 4 (i.e. equivalent to B2) for Matric SSC I and level 5 (C1) for SSC II. There is, indeed, a progression across the 4 levels of the CEFR from A2-C1 for Matric SSC II but for SSC I the sample could only progress through 3 levels (A2-B2). Level A1 could not be awarded as all the sample papers were able to write more than “simple isolated phrases and sentences” as the descriptor for over all written production suggests (Council of Europe, 2001:61). The sample performances could not provide sufficient evidence to be awarded at higher CEFR levels i.e. C2 level. Only on one instance C1 was awarded to SSC II sample performance. However, the scores awarded by the teachers represented 4 performances (2 SSC I and 2 SSC II) at C2 levels.

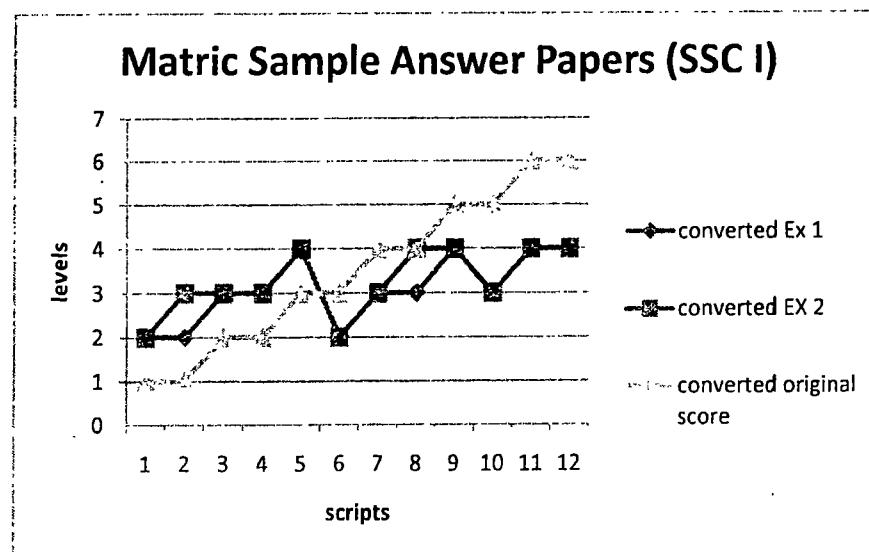


Figure 14: Summary of rating Sample Matric English Answer Papers (SSC I)

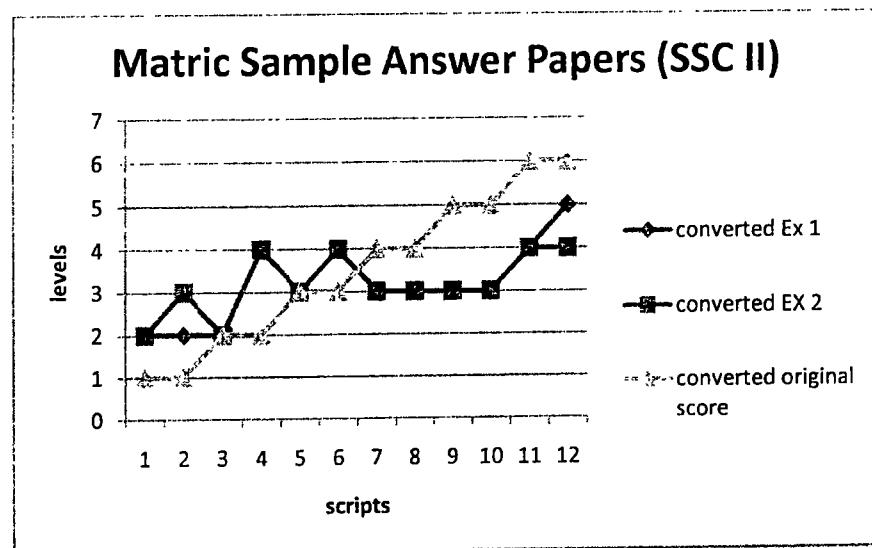


Figure 15: Summary of rating Sample Matric English Answer Papers (SSC II)

The examiners generally agreed that the Writing skills required and elicited in Matric English exam can only be judged as being at the level of A2-B2 levels of CEFR. However, these ratings should be interpreted with caution because as already discussed the focus was on the writing tasks rather than the reading ones as they were based on the text books and did not require the candidates to use their Reading skills. The Reading skills then required and elicited in Matric English exam can only be judged as being at the level of Pre A1-B1 levels of CEFR. Pre A1 because CEFR levels do not consider reproducing memorized responses to predicted question types based on the text book.

5.2.3 Validation

The results show that the expected Matric English production levels (i.e. Writing) are higher than the Reading Comprehension levels keeping in view that the reading sections of the paper are based on the text book. From the observations and discussions throughout the alignment process, it was felt that the claimed Matric English Reading and Writing levels as

by IBCC are higher than the actual English question papers and performance levels of the test takers.

To conclude, the alignment of Matric to CEFR can be verified informally. Although, a more confident conclusion could have been drawn in the presence of the required number of experts and the Matric sample answer papers from Annual board examinations and marked by the examiners under FBISE rules and regulations. It was, however, an awareness-raising exercise as the participants were able to highlight the gaps between the Matric English testing procedures and serves as a preliminary study to the actual alignment activity according to the standard procedures of the Manual (2009). Although, after the provision of all the evidence required for the quality of exam and the appropriate number of panel of experts, the alignment would be a useful exercise to promote good testing practices and therefore improve quality of examinations in Pakistan.

5.2.4 Some Procedural Issues

Besides the issues of acquiring the required local sample answer papers from FBISE, there was one issue which was felt very important related to consulting different CEFR tables. As also highlighted in Salamoura and Papp (2009), regarding the use of materials in the linking exercise, consulting too many tables for each skill was a little confusing. The more desirable and convenient way would be to collate all the tables for each individual skill into one detailed table e.g. Reading, Writing, Speaking, Listening etc.

CHAPTER 6

FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

This chapter gives a recap of the findings and possible solutions. It suggests scope of the future research studies basing on the evidence derived from this study. Findings are discussed along with some recommendations and suggestions for the English curriculum designers, teachers and test developers. Further suggestions to update the existing Matric syllabus and adapt the content and teaching/assessment methodology and bring it at par to any internationally recognized system will be included in this chapter.

6.1 Findings

To review, the first research hypothesis is accepted that, “The conversion scale provided by IBCC of converting scores from O Level to Matric undervalues the performance of O Level students by comparison with Matric students,” it was found that the current conversion scale of IBCC undervalues the performance of O Level students by comparison with Matric students and needs to be revised. The evidence provided by the current findings suggests that the current conversion ratings of O Level grades in English are not appropriate, and would suggest that their valuations vis-a-vis the Matric scale should be increased by at least 5 points. This undervaluation of O Level grades might be one of the key reasons and rationale to prefer Matric over O Level.

Looking at the second and the third hypotheses,

“Students of O Levels score better on a Common English Test of reading and writing”,

“Matric English testing system if compared to English testing system of O Level is not equivalent,”

the study indicates that on average the O Level students clearly out performed Matric students on the common reading and writing test. Therefore, it was concluded that in light of the present findings of the current study, O Level students seem better in terms of their English reading and writing skills on average than the Matric students. Thus, considering the findings from the first phase of the study both the hypotheses are accepted.

Reviewing the fourth hypothesis,

“English test scores from O Level are more valid and reliable than Matric,”

it is concluded that on the grounds that O Level is internationally recognized and is aligned with the CEFR, its scores certainly have more validity and reliability than the scores from Matric. Moreover, in this study too, as the O Level scores are more compatible with scores from the Common English Reading and Writing Test, so it makes them more valid and reliable. Therefore, the evidence provided in this study suggests to accept the fourth hypothesis.

Looking at the fifth hypothesis,

“Matric English testing procedures practiced by FBISE are not testing the same level of

English reading and writing skills of students as O Level by CIE, Cambridge,”

it was found that in the present circumstances, in the absence of the validity and reliability proof of Matric English papers, the in depth analysis of the papers in the Phase 2 of the study resulted in the conclusion that the English exam tasks were meant for B1 CEFR level mostly, except for a few B2-C1 level Writing tasks. Reading tasks based on the text book requiring

memory skills could be considered pre-A1 level while some reading tasks required and elicited responses of B1 level. In other words, O Level tests at levels of C1/C2 of CEFR and Matric being its alternative in Pakistan should also do the same and improve the level so the test takers strive for a higher level of CEFR that is the higher communicative competence in English language. The sample Writing performances on the other hand, suggest that they were able to elicit answers up to C1 level. Therefore, the hypothesis is accepted.

In answer to the sixth hypothesis,

“Matric English testing procedures practiced by FBISE do not possess the same level of validity as O Level when evaluated through the alignment process of CEFR, an international framework, to which O Level already is aligned.”

it was found through the content analysis of Matric English testing procedures practiced by FBISE do not possess the same level of validity as O Level when evaluated through the alignment process of CEFR. Furthermore, the alignment process to CEFR is possible with a few critical issues to be resolved like the proof of validity of the test. In the second phase where the alignment process was carried out, it was concluded that having completed all the three major steps of the alignment process according to all the resources available, the task is possible. As mentioned by Szabo (2010), linking itself is never completed, actually it only particular tasks that can at certain points be considered linked. So, indeed this preliminary study provided a base for the possibility of a more extensive and complete alignment process to a never complete linking process. It was proved, however, that Matric English testing procedures practiced by FBISE do not possess the same level of validity as O Level, for instance, there is little evidence of ensuring the reliability and validity of Matric English exam paper content. The level of the tasks in the question papers do not go beyond B2 of

CEFR unlike O Level. Furthermore, most of the question types are memory based and not skilled based.

Moreover, to keep abreast with the changing needs and trends of the world, there is a need to formulate assessment groups like in the more developed parts of the world to harness the education system with controlled and monitored guidance. There are many researchers ready to help even on voluntary basis only if there is assurance of their efforts not to go in vain. The largest assessment reform group in the UK such as Assessment Reform Group was a voluntary group of researchers brought together by British Educational Research Association (BERA). There is no dearth of such dedicated volunteer researchers in Pakistan ready to make a mark in the educational field but they need an anchoring point.

The need of evolution is most felt now than ever before and Pakistan is ready for a change towards better testing practices. As Harrison (1991) points out that change takes time. He documents few cases e.g. in the case of the Institute of Linguists' examinations, it took ten years in the West Bengal School examination's case. Therefore, today's effort will bear fruit in probably a decade.

6.2 Conclusions

This study has tried to develop a methodology and a way to address a practical and long felt need for the comparability of education standards throughout Pakistan. Although, all the boards could not be taken into account but addressing the issue by focusing one of the largest educational board of Pakistan (i.e. FBISE see section 3.3.1) is a good starting point. Studies addressing the comparability of Secondary English examinations has until now been lacking in Pakistan. Although, there is no absolute universal method of analysing different examinations but an attempt has been made to propose and demonstrate one way of carrying out such a task in a principled fashion.

According to the findings of this study, it has been empirically substantiated that the O Level group clearly out performs the Matric group in the English Reading and Writing Common Test. O Level group out performs Matric group by one point on the Common Test (subtest of IELTS) which is significant on the scale of 9 points. This is an important finding because, in fact, even half a point can restrict a candidate's academic and professional prospects.

The question arises if the sample does not differ in their demographic conditions and come from the same schools, then the difference in their English language competence certainly has something to do with the system that they follow. It has been proven through this study that O Level equips the students better with English language skills than Matric.

6.3 Recommendations: Areas for Follow-up and Further Research

There are various issues raised in this research which could benefit from further research. It will be an exaggeration to say that the methodology adopted in this study was the only way to improve the existing testing system in Pakistan. There are considerable amount of validation procedures found in the literature which have demonstrated remarkable results in providing results for the validation of examinations (e.g. Alderson and Buck, 1993; Hawkey, 2004; Bachman, 1990; 1996; 2004 among others). The results from this present study highlight some important issues and pose some important questions for future research to improve the existing Matric English testing system.

It is clear that one follow up needed is the successful completion of the alignment process as per CEFR Manual (2009). This study has provided a preliminary study which can be taken a step ahead to align Matric to CEFR properly with all the required resources. The impact of linking to CEFR is manifold. First, it will further make Matric transparent by revealing how comparable it is to any international framework of reference highlighting its

shortcomings and areas to be improved to be at par to any international test. Second, CEFR having the potential to be used in Pakistani context will give familiarity and awareness of CEFR levels to the masses and how the proficiency levels work around the globe.

For providing high quality language tests, linking is worth doing in order to connect the relevant test users in Pakistan to the international testing community and the candidates.

A second area for follow up would be revising the IBCC scales for the conversion of O Level scores into Matric. It would be interesting and very beneficial to conduct a similar research study covering a broader range of sample including all the other parts of Pakistan putting all the secondary education boards to test. One might hazard a guess that the situation in other secondary educational boards will not be very different from FBISE.

A third follow up is to take all the secondary educational boards and compare them in terms of their examination procedures, examination content in order to equate all these boards and set standard procedures to ensure quality. This will help in maintaining the standards of testing and indirectly the teaching.

A fourth follow up would be to use CEFR (2001) to the utmost e.g. besides using it for designing the curricula and syllabi, it can also be used for improved teaching and lesson planning in an English classroom. CEFR (2001) provides comprehensive guidelines for English language teachers.

A final and most significant follow up would be to try to formulate procedures to capture the areas that this study failed to address adequately.

The very final and the foremost follow up would be to promote good testing practices in general at all levels. Similar research studies raising questions and exploring the issues will bear fruit and a positive change might emerge on the scene of English language testing in Pakistan. Results from the present study are certainly encouraging and a hope to stream

line things towards a more transparent and coherent examination system and thus a better educational system.

There would be those who would argue that the present system is the best and the most suitable system in the current resources available. They might resist any efforts to change the system but where there is a will there is a way. Every year there is a specific fund in the Annual budget of Pakistan reserved for the education sector. There is need to use the available funds for education wisely and appropriately in order to improve not only the testing system but the teaching practices and then maintain the standards for quality education. FBISE fails to deliver testing procedures of international standard that it claims to, which is evident when compared to another internationally recognized system. In theory, such a system could be adopted in Pakistan but in practice such a task would be difficult indeed but not impossible to achieve. Pakistan needs an educational reform; we need to evaluate our education policies and proposals for change.

CHAPTER 7

IMPLICATIONS FOR MATRIC ENGLISH TEST AND TESTING PROCEDURES

This last chapter discusses the various aspects of an updated Matric English test and testing procedures to promote good testing practices in Pakistan.

7.1 A Uniform Change

Uniform change across all the educational boards is the only solution. As can be deduced from FBISE's effort to shift from memory to skill based examination when in 2005, it tried to adapt a more communicative approach to language testing with the help of an experienced team from Edexcel, UK. An effective change could only be achieved if all the educational boards across the country are compelled to implement the change simultaneously. The process could begin with implementation of the already developed English curriculum based on communicative approach and extending the same approach to the testing procedures (Ministry of Education, 2006).

7.2 Alignment to CEFR

The foremost need for an updated and better Matric would be its alignment to CEFR. The shortcomings of Matric English testing system that emerged in this preliminary study can be studied in depth with properly carrying out the alignment process.

7.3 Communicative Approach to Teaching and Testing

The English curriculum (Ministry of Education, 2006) certainly is totally based on the communicative approach (see Chapter 3, section 3.11.4) and is quite comprehensive but the pedagogical practices in a language classroom are still clinging on to the Grammar Translation Method. Most teaching in a typical Pakistani classroom is usually focused on fact transfer and information recall which is the lowest order of Cognitive skill according to Bloom's Taxonomy (1956) rather than true meaningful, intellectual and high order Cognitive skill development. This promotes rote learning rather than skill learning. This remains a central challenge for educators and trainers in modern times. Vernon (1956) also asserts that cramming and rote memorization generally stuns the reasoning power of the pupil.

The reasons are large classrooms, lack of teacher training, lack of resources. But that does not mean standards cannot be improved.

CEFR (2001) which emphasizes learners' communicative needs should be used to the utmost e.g. besides using it for designing the curricula and syllabi, it should also be used for improved teaching and lesson planning in an English classroom.

7.4 An Aligned Curriculum

The English curriculum of Matric as described in Chapter 3 (section 3.11.4), is based on communicative approach proposing pedagogic tasks in a language classroom promoting the communicative competence of the pupil. These authentic tasks, according to CALT, are important for promoting relevant and challenging exchanges which in turn encourage meaningful communications. So, like any other effective scheme of studies Matric also proposes clear learning outcomes. As Biggs (1999) puts it, effective teaching is supported by aligned curriculum which apart from clear learning outcomes requires learning experiences designed to assist learners achievement of those out comes. Matric English objectives are

specified in terms of cognitive skills which are generally derived from the Cognitive Domain of Bloom's Taxonomy (1956).

Curriculum and test designers' generally develop course objectives and design tests within some framework of learning, teaching and assessment. One of the most popular ones is Bloom's (1956) "Taxonomy of Learning Domains" or "Taxonomy of Educational Objectives", translated in over 22 languages (Anderson & Sosniak, 1994). This Taxonomy claims to promote true meaningful intellectual and skill development in a classroom. Although, it is a challenging task but Bloom's Taxonomy can be used which provides an excellent structure for planning, designing, assessing and evaluating teaching and learning effectiveness. The framework also serves as a sort of checklist, by which the teacher and the assessor can ensure that teaching is planned to deliver all the necessary skill development for students, trainees or learners, and a template by which the validity of any test can be assessed.

An aligned curriculum, according to Biggs (1999), also needs carefully developed assessment tasks that allow students to demonstrate achievement of the learning outcomes sketc.hed out at the beginning of the course so they are in fact a linkage between the expectations, teaching and assessment. For Hughes (1989) such achievement tests which are based on the course objectives are preferred rather than those based on the detailed course content. The advantage that he outlines is that such tests offer a clear picture of how far the objectives set at the beginning of the course have been achieved. This obviously puts pressure on those who are responsible for the syllabus and selection of the text books and materials to ensure that they are consistent with the course objectives. Kellaghan and Greaney (1992) also insisted that the examination should be a reflection of the entire

curriculum and not merely a specific part of it, thus propagating the need of an aligned curriculum,

The English curriculum of Matric being as good as the O Level curriculum does not seem to bear any pressure on the text book writers and the books used in both the systems have poor compatibility or the lack of maintaining the standards. Matric uses very simple, easy English text books, one for class 9 and the other for class 10 and these books are the only resources used because most of the paper is based on these books, therefore promoting rote learning or in Eisemon (1990) words “bad cramming”. In such a situation the teachers tend to focus more on the examination objectives rather than on the syllabus objectives and tailor their teaching methodologies with an aim to help the student pass the examinations as the stakes of this test are very high. While O Level uses many resources or the students and teachers cannot rely on only one source as the paper content is unseen and if emphasis be on any one source can seriously jeopardize the success rate of the students in the examinations.

As Davies (1977) points out that language testing influence the curriculum and teaching so it is important that the language test should be well constructed. Therefore, for a valid and reliable test it is of prime importance for its content to be in line with the curriculum objectives but the English papers of Matric are not in line with its curriculum objectives. The curriculum clearly states:

“The focus of the curriculum is to prepare students for an assessment of their acquisition and use of language skill rather than memorization of the text book contents. Discouraging rote learning will enable students to approach different texts independently. Skill-based assessment using unseen texts and materials is, therefore, recommended for testing students’ ability to use language in spoken and written communication” (Ministry of Education, Pakistan, 2006: 153)

This objective does not seem to be materialized in the Matric English test content. As Wiseman (1961) claims that in order to avoid the negative impact of examinations, it is

necessary to keep aims of curriculum in mind while setting papers. Thus, there is a need for the Matric English course objectives set at the beginning of the course to be in line with the test items at the end of academic year in the terminal examinations. Hence, following Kellaghan and Greaney (1992) guidelines for an improved public examination system the educational authorities, curriculum developers, test writers would all collaborate to have a harmonized educational system and work as a team.

7.5 Validation of Examination System

Validation of Matric is the most important issue which if catered will solve most of the problems. Keeping in view Kellaghan and Greaney's (1992) suggestions, it is important for improved and standardized examination system to have the capacity to conduct research by each of the examination board. One of the most robust checklist proposed by Alderson and Buck's (1993) posing important questions on tests questioning their validity. The criteria for validation processes and aspects proposed by Alderson and Buck (1993) adapted into a table by Hawkey, (2004:31, chapter 3) if applied to Matric English Test will crease out many of its shortcomings.

7.5.1 Concurrent and Predictive Validity. Following Kellaghan and Greaney's (1992) recommendations when they stress the importance of each examination board having the capacity to conduct research, they believe that predictive validity studies of public examinations specifically should be conducted. So, there should be test validity studies for equivalence of different versions and forms of Matric English test. Studies providing quantitative or qualitative analyses of the test make a test more transparent and valid.

7.5.2 Construct and Content. Assessment objectives should be specified and as Pollitt et.al (2007) notes that it is a common practice for O Level to specify the content of papers in terms of cognitive skills which are generally derived from the cognitive domain of

Bloom's taxonomy (1956), so Matric should follow the same. Curriculum and test designers' should develop course objectives and design tests within the framework provided by Bloom and include the higher order cognitive skills in the tests especially as Kellaghan and Greaney (1992) recommends that this practice will ensure that the higher order cognitive skills are also taught in the classrooms.

There is, however, a debate amongst the test writers whether the test items should be based on the course objectives or the detailed course syllabus or on the books and other materials used in the classroom. Hughes (1989) prefers that course-objective tests rather than the course-content-based tests as they are more effective. Matric system might be using the latter approach of basing the tests on the books and other materials used in the classroom or rather a single text book but the FBISE staff does not admit it explicitly as interviewed by the researcher⁵¹.

Information on content of exam should be clearly defined along with the purpose of the examination keeping the target candidates in mind. An appropriate difficulty level for concerned level should be dealt with sensitively. The issue of validity again deals with the alignment of the course objectives to the test items. Unless the test measures what it is meant to measure, it cannot claim to have validity and thus the inferences drawn from such a test cannot be reliable. As seen in the alignment to CEFR process, the Matric papers are mostly based on the memory i.e. requiring the candidates to recall from their memory which for any age group is very invalid but particularly at this age (i.e. 14-16 years old) is quite inappropriate as Vernon (1956) believes that it stuns the intellectual development of the pupil. The other tasks even which required something beyond memory were also not suitable for the age group in question and the desired level e.g. most of the tasks only required level

⁵¹ In the meeting with directors of the curriculum and examination wing of Federal Board of Intermediate and Secondary Education (FBISE), they claimed their tests were based on the course objectives rather than on the detailed course content.

B1-B2 of the CEFR levels which is quite basic and not appropriate for a secondary school leaving candidate especially when the language of instruction in colleges is English. All this information should be accessible to the students. There is a dire need of systematic needs analyses of key stakeholders. All this additional information should be provided to the item writers to cater for all the stakeholders.

7.5.3 Exam Construction Process. All the components that goes into the examination construction e.g. Item writers, item writing, moderating has to be evaluated and revised. Pre-testing, yet another important link in the examination construction which needs considerable attention. The training of item writers, editing and checking of the items need to be improved. Statistical analyses of pre-tests need to be implemented to prove empirically the worth of the test items.

The curriculum designers first need to make sure they design a course involving all the 6 components of Cognitive Domain of Bloom's Taxonomy taking the student from the lowest to the highest level to master a specific skill. Testing specialists, on the other hand, require measuring the students' skill and identifying his/her level of competence and for that they need to analyse whether the test items are in line with the course objectives. It is of utmost importance for a reliable and valid measurement to ensure the alignment of the course and assessment objectives. Including more low levels of Bloom's Cognitive paradigm and ignoring the higher ones would yield imperfect measurement putting the validity of the test content and scores in question. So, a valid and reliable test would include questions comprising different levels of difficulty to test the different levels of students' capabilities.

The examination papers should be developed along with their marking schemes based on CEFR. The papers should be designed by the examiners who are experienced and have

had some formal training and apprenticeship. The exam content should be based on the curriculum; testing the skills, knowledge and understanding of the candidates.

Post hoc Analysis and Reporting should be a necessary component of a testing process. Statistical analyses of exams followed by exam reports should be provided as their main responsibility. All these analyses and reports should be accessible to all the relevant test users.

Exams should be revised in the light of exam feedback from the experts and the relevant test users incorporating the modern and current learning theories. Student feedback should be given due importance and systematic revision procedures should be devised. However, frequency of exam revisions should have a valid rationale taking all the test users in account.

7.6 Reliability

Successful administration of a test is another important link in the examination process. Responsibility for administering the exams, training and monitoring are vital features that have to be handled with care and proper methods.

As mentioned earlier, reliability of the markers had been questioned for centuries and has always been the focus of testing research. Standardization procedures should be undertaken. There should be measures to estimate rater reliability and consistency by the statistical measures which should be a routine practice for any high-stakes public exam. There has to be ways to determine the consistency of results.

There should be liberty to challenge the awarded grade and request for reassessment if the student believes that he/she has not been awarded a fair and correct grade. The request for reassessment should be the right available to all the students. The idea of an independent body as in the UK i.e. the Examinations Appeals Board should also be looked into for a fair

and just testing system (see Chapter 3, section 3.10.5.6). The need for Matric is to follow such a mechanism which has already been developed by the developed West, the effort is only to reach out to the fair and transparent system and implement it honestly.

7.7 Examiners Training

Training of the examiners is yet another issue which needs to be properly addressed. There should be training for the examiners/ markers for a standardized marking. The examiners should be trained through marking sample scripts and discuss assessment criteria thoroughly with other examiners and head examiner. If all the procedures set out in the CEFR Manual (2009) are followed, the problem could be solved with concrete outcomes. The positive impact of such innovations would be felt not only in the educational system but the whole society.

7.8 Marking Scheme of Matric English Papers

The marking key provided to the examiners should be based on analytical marking. The testees should have a clear idea how they will be assessed and which areas hold more weightage than the others and so on with clear grade descriptors. Comprehensive detail about the specification of examinations, the syllabus aims and objectives, the assessment aims and objectives should be accessible to all the testees. The descriptive marking rating criteria for each paper, previous papers along with their detailed marking keys should also be easily accessible. Thus, the testees should have a very fair idea about what and how they will be marked making the exam transparent and reliable.

References

Alderson, J. C. (1986). Innovations in Language Testing? In M. Portal (ed.) *Innovations in Language Testing*. London: NEFR/ Nelson, 93-105.

Alderson, J. C. (1991). Language Testing in the 1990s: How Far have We Come? How Much Farther have We to Go? In S. Anivan (ed.) *Current Developments in Language Testing*. Singapore: SEAMEO Regional Centre.

Alderson, J. C. & Buck, G. (1993). Standards in Testing: A Study of The Practice Of UK Examination Boards in EFL/ESL Testing. *Language Testing*, 10, 1, 1-26.

Alderson, J. C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, and A. Curtis (eds.) *Washback in Language Testing: Research Contexts and Methods*. London: Lawrence Erlbaum.

Alderson, J. C. & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14, 2, 115 - 129.

Anderson, L. W., & Krathwohl, D. R. (eds.) (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision Of Bloom's Taxonomy of Educational Objectives: Complete Edition*. New York: Longman.

Anderson, L. W. & Sosniak, L. A. (eds.) (1994). *Bloom's Taxonomy: A Forty-Year Perspective*. *Ninety-third Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.

Assessment Reform Group (ARG) (2000). *Beyond Black Box*. Cambridge: Faculty of Education, University of Cambridge.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L.F. (2004). *Statistical Analysis for Language Assessment*. Cambridge: Cambridge University Press.

Bachman, L.F. & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Bachman, L.F., Davidson, F., Ryan, K. & Choi, I.C. (1995). *An Investigation into the comparability of two tests of English as a foreign Language: the Cambridge-TOEFL comparability study*. Cambridge: University of Local Examinations Syndicate and Cambridge University Press.

Bailey, K. (1996). Working for Washback: A Review of the Washback Concept in Language Testing. *Language Testing*, 13 (3), 257-279.

Barni, M. Scaglioso, A. M. & Machetti, S. (2010). Linking the CILS Examinations to the CEFR. In M. Milanovic and C. J. Weir (eds.) *Aligning Tests with the CEFR*. Vol 33, Cambridge: UCLES/Cambridge University Press, 159-176.

Baker, D. (1989). *Language Testing: A Critical Survey and Practical Guide*. London: Edward Arnold.

Bott Van Houten, J. (2007). *Language learning policies in the United States*. Paper presented at the Council of Europe's Intergovernmental Policy Forum: The Common European Framework of Reference for Languages (CEFR) and the Development of Language Policies: Challenges and Responsibilities, Strasbourg, France.

Biggs, J. B. (1995). Assumptions Underlying New Approaches to Educational Assessment. *Curriculum Forum*, 4 (2), 1-22.

Biggs, J. (1999). *Teaching for Quality Learning at University*. Society for Research into Higher Education and Open University Press, Oxford.

Bloom B. S. (1956). *Taxonomy of Educational Objectives*. Handbook I: The Cognitive Domain. New York: David McKay.

Borsboom, D. & Mellenbergh, G.J. (2004). Why Psychometrics is not Pathological. *Theory and Psychology*, 14, 105-120.

Brannen, J. (1992). Combining Qualitative and Quantitative Approaches: An Overview. In J. Brannen (ed.) *Mixing Methods: Qualitative and Quantitative Research*. Burlington, VT: Ashgate, 3-37.

Brown, J. D. (2003). Creating a Complete Language-Testing Program. In C. A. Coombe & N. J. Hubley (eds.) *Assessment Practices*. Washington, DC: TESOL, 9-23.

Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.

Cambridge ESOL (2011). *Using the CEFR: Principles of Good Practice*. Available online at www.CambridgeESOL.org

Canadian Language Benchmarks (2006). *Companion Tables to the Canadian Language Benchmarks 2000*. Canada:Centre for Canadian Language Benchmarks. Available online at http://www.language.ca/cclb_files/doc_viewer_dex.asp?doc_id=247&page_id=50. Retrieved 22.9.2012.

Canale, M. & M. Swain (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1) 1-47.

Caroll, J.B. (1991). Resistance to Change. In C. Alderson & B. North (eds). *Language Testing in the 1990s*. GB: Macmillan Publishers Limited, 22-27.

Cheng, L. (2005). Changing Language Teaching Through Language Testing: A Washback Study. In M. Milanovic & C. J. Weir (eds.) *Studies in Language Testing Vol. 21* Cambridge: Cambridge University Press.

Choi, I. C., Kim, K.S. & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20 (3), 295-320.

Coe, R. (2007). Common examinee methods. In P. Newton , J. Baird, H. Goldstein, H. Patrick, & P. Tymms (eds.) *Techniques for Monitoring the Comparability of Examination Standards*. Malta, 331-367.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Council of Europe (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A*

Manual. Strasbourg: Council of Europe Language Policy Division. Available online at http://www.coe.int/t/dg4/linguistic/Manuell_EN.asp#TopOfPage. Retrieved 12.12.2012.

Cloud, N., Genesse, F. & Hamayan, E. (2000). *Dual Language Instruction: A Handbook for Enriched Education*. Boston: Heinle & Heinle.

Crawford, J. (2004). *Educating English learners*. Los Angeles: Bilingual Education Services.

Davies, A. (ed.) (1968). *Language Testing Symposium: A Psycholinguistic Approach*. London: Oxford University Press.

Davies, A. (1977). The Construction of Language Tests. In J. B. Allen and A. Davies (eds) *Testing and Experimental Methods*, Vol. 4 in *The Edinburgh Course in Applied Linguistics*. London: Oxford University Press, 38-104.

Davies, A. (1978). Language Testing. In: *Language Teaching and Linguistics Abstracts*. Vol. 11, 3 & 4, reprinted in V. Kinsella (ed.) (1982) *Surveys 1: Eight State-of-the-art Articles on Key Areas in Language Teaching*. Cambridge University Press, Cambridge, 127-159.

Davies, A. (1985). Following Leader: is what Language tests do? In Y. P. Lee, A.C.Y.Y. Folk, R. Lord & G. Low (eds.) *New Directions in Language Teaching*. Cambridge: Oxford: Pergamon Press, 3-12.

Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

Downey, N. & Kollias, C. (2010). Mapping the Advanced Level Certificate in English onto the CEFR. In M. Milanovic and C. J. Weir (eds.) *Aligning Tests with the CEFR*. Vol 33, Cambridge: UCLES/Cambridge University Press, 119-129.

Eckes, T., Ellis, M., Kalnberzina, V., Pizorn, K., Springer, C., Szollas, K. & Tsagari, C. (2005). Progress and Problem in Reforming Public Language Examinations in Europe: Cameos from the Baltic States, Greece, Hungary, Poland, Slovenia, France, and Germany. *Language Testing*, 22, 355-377.

Eckstein, M. A. & Noah, H. J. (1993). *Secondary School Examinations; International Perspectives on Policies and Practice*. New Haven: Yale University Press.

Edgeworth, F.Y. (1888). The Statistics of Examinations. *Journal of the Royal Statistical Society*, 51, 599-635.

Eisemon (1990), T O (1990) Examinations Policies to Strengthen Primary Schooling in African Countries, *International Journal of Educational Development* 10(1), 69-82.

Forehand, M. (2005). Bloom's Taxonomy: Original and Revised. In M. Orey (ed.) *Emerging Perspectives on Learning, Teaching, and Technology*. Available online at <http://projects.coe.uga.edu/epltt/> Retrieved 21.3.2012.

Foster, P. J. (1992). Commentary, Chapter 9 in M. A. Eckstein & H. J. Noah (eds) *Examinations: Comparative and International Studies*. Oxford: Pergamon Press.

Frederiksen, J. R. & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher*, 18 (9), 27-32.

Freeman, Y., Freeman, D. & Mercuri, S. (2005). *Dual Language Essentials for Teachers and Administrators*. Portsmouth, NH: Heinemann.

Fulcher, G. (2000). The Communicative Legacy in Language Testing. *System*, 28, 483–497.

Fulcher, G. (2007). Universities undermine their own foundations. In *The Guardian Weekly TEFL Supplement*, 13.

Fulcher, G. (2010). *Practical Language Testing*. UK; Hachette UK Company.

Gardner, J. et. al. (2008). *Changing Assessment Practice: Process, Principles and Standards*.

Available online at

[http://www.nuffieldfoundation.org/sites/default/files/JG%20Changing%20Assessment%20Practice%20Final%20Final\(1\)..pdf](http://www.nuffieldfoundation.org/sites/default/files/JG%20Changing%20Assessment%20Practice%20Final%20Final(1)..pdf) Retrieved 23.2.2012

Genesee, F. (1994). Assessment Alternatives. *TESOL Matters*, 4 (5), 2.

Goullier, F. (2007). *Council of Europe Tools for Language Teaching: Common European Framework and Portfolios*. Available online at

http://www.google.com.pk/ur!sa=t&rct=j&q=goullier%202007&source=web&cd=3&cad=rja&ved=0CDIQFjAC&url=http%3A%2F%2Fwww.coe.int%2F%2Fdg4%2Flinguistic%2FSource%2FGoullier_Outils_EN.pdf&ei=Jbr3UYnuHleRO5CWgYgN&usg=AFQjCNFcXKUm7BDNn7cN2MvEcPzZ3wlimQ&bvm=bv.49967636,d.Yms Retrieved

23.9.2012

Gupta, A. (2007). *Going to School in South Asia*. USA: Greenwood Press.

Hamp-Lyons, L. (1997). Washback, Impact and Validity: Ethical Concerns. *Language Testing*, 14 (3), 295–303.

Halliday, M. (1976). The Form of a Functional Grammar. In G. Kress (ed). *Halliday: System and Function in Language*. Oxford: Oxford University Press.

Harrison, A. (1983). *A Language Testing Handbook*. London: Macmillan Press.

Hawkey, R. (2004). *A Modular Approach to Testing English Language Skills, Studies in Language Testing*, Vol 16, Cambridge: UCLES/Cambridge University Press.

Hughes, A. (1989/2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Harrison, A. (1991). Language Assessment as Theatre; Ten Years of Communicative Testing. In C. Alderson & B. North (eds.) *Language Testing in the 1990s*. GB: Macmillan Publishers Limited, 95-105.

Hymes, D. (1972). On Communicative Competence. In J. Pride and J. Holmes (eds.) *Sociolinguistics*. Harmondsworth: Penguin, 269-293.

Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38 (4) 319 - 342.

Kellaghan, T. & Greaney, V. (1992). *Using Examinations to Improve Education: A Study of Fourteen African Countries*. Washington, D.C: The World Bank.

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. New York: MacMillan.

Kirkland, M. C. (1971). The Effect of Tests on Students and Schools. *Review of Educational Research*, 41, 4, 303-350.

Kluitmann, S. (2008). *Testing English as a Foreign Language: Two EFL-Tests Used in Germany*. MA Thesis, University of Albert-Ludwig.

Kostopoulou, S. (2008). Democratising and Enhancing the Quality of Institutionalized Language Assessment through the European language Portfolio. In L. Taylor & C. J. Weir (eds.) *Language Testing Matter*. Cambridge: Cambridge University Press.

Kunnan, A. J. (1999). Fundamentals [in language testing]. In B. Spolsky (ed.) *The Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier Science, 707-711.

Linacre, J.M. (1991). *Structured rating scales*. Sixth International Objective Measurement Workshop. April. Chicago. ERIC TM 016615

Linn, R. L. (1983). Teaching and Instruction; Links and Distinctions. *Journal of Educational Measurement*, 20, 179-189.

Linn, R. L. (1992). *Educational Assessment: Expanded Expectations and Challenges*. CSE Technical Report 351. Boulder: University of Colorado at Boulder, CRESST.

Lumley, T. & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12 (1) 54-71.

Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.

Madaus, G. F. (1988). The Influence of Testing on the Curriculum. In L. N. Turner (ed.) *Critical Issues in Curriculum: Eighty-seventh Yearbook of the National Society for the Study of Education*. Chicago: NSSE, University of Chicago Press, 83-121.

Mansoor, S. (2005). *Language Planning in Higher Education: A Case Study of Pakistan*. Karachi: Oxford University Press.

Martyniuk, W. (2008). Relating language examinations to the Council of Europe's Common European Framework of Reference for Languages (CEFR). In L. Taylor & C. J. Weir (eds.) *Multilingualism and Assessment*. Cambridge: Cambridge University Press, 9-20.

Milanovic, M. & Weir, C.J. (2010). *Aligning Tests with the CEFR*. Vol 33, Cambridge: UCLES/Cambridge University Press.

Mathews, J. C. (1985). *Examinations: A Commentary*. London: George Allen and Unwin.

McMillan, J. H. (1996). *Educational research: Fundamentals for the Consumer*, second edition. New York: Harper Collins College Publishers.

McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.

Messick, S. (1989). Validity. In R. L. Linn (ed.) *Educational Measurement*. New York: American Council on Education/ Macmillan, 13-103.

Messick, S. (1996). Validity and Washback in Language Testing. *Language Testing*, 13 (3) 241-256.

Milanovic, M. & Saville, N. (1996). *Considering the Impact of Cambridge EFL Examinations*. Internal Working Report. Cambridge: University of Cambridge Local Examinations Syndicate.

Ministry of Education (Pakistan). (2006). National Curriculum for English Language Grades I- XII. Available online at

<http://www.sharepdf.com/a79b0aa483ca43669f0109d7ad538c9e/pdf%20f.pdf>

Retrieved 1.5.2014.

Morrow, K. (1977). *Techniques of Evaluation for a National Syllabus*. Study commissioned for the Royal Society of Arts. Reading: Centre for Applied Language Studies.

Morrow, K. (1979). Communicative Language Testing: Revolution or Evolution? In C.K. Brumfit, and K. Johnson (eds.) *The Communicative Approach to Language Teaching*. Oxford: Oxford University Press, 143–159.

Morrow, K. (1986). The Evaluation of Tests of Communication Performance. In M. Portal (ed.) *Innovations in Language Testing*. London: NEFR/Nelson.

Murphy, R. (1982), Sex Differences in Objective Test Performance. *British Journal of Educational Psychology*, 52, 213–219.

Murphy, R. (2007). Common Test methods. In P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (eds.) *Techniques for Monitoring the Comparability of Examination Standards*. Malta, 301-323.

Naeem, M. I. (2011). *A Comparative Study of Secondary School Certificate (SSC) and of General Certificate of Education- Ordinary Level (GCE-OLevel)*. Unpublished Manuscript. Pakistan: International Islamic University.

Newton, P., Baird, J., Goldstein, H., Patrick, H., & Tymms, P. (2007). *Techniques for Monitoring the Comparability of Examination Standards*. Malta.

Noah, H. J. & Eckstein, M. A. (1989). Tradeoffs in Examination Policies: An International Comparative Perspective. *Oxford Review of Education*, 15 (1) 17-27.

Nordic Recognition Information Centres (2006). *Report on the system of Education in Pakistan*. Available online at http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Kunnskapsbasen/Konferanser/SU%20konferanser/Seminarer/Fagseminar_06/The%20System%20of%20Education%20in%20Pakistan.pdf Retrieved 12.9.2012.

Noble, A. J. & Smith, M. L. (1994). *Old and New Beliefs about Measurement Driven Reform: The More Things Change, The More they Stay the Same*. CSE Technical Report 373, AZ: Arizona State University, CSE.

North, B. (1992). Options for Scales of Proficiency for a European Language Framework. In R. Schiirer & B. North (eds.) *Toward a Common European Framework for Reporting Language Competency*. Washington, DC: National Foreign Language Center, 9-26.

North, B. (2000). *The Development of a Common Reference Scale of Language Proficiency*. New York: Peter Lang.

North, B. (2007). *The CEFR levels: Key Points and Key Problems*. Paper presented at the 23rd ALTE Conference, Sevres.

North, B. (2009) The Educational and Social Impact of the CEFR. In L.Taylor & C.J. Weir (eds.) *Language Testing Matters, Investigating the Wider Social and Educational Impact of Assessment*. Proceedings of the ALTE Cambridge Conference, April 2008, Studies in Language Testing, 31 Cambridge: Cambridge University Press, 357-377

Patrick, H. (1996). *Comparing Public Examination Standards over Time*. Paper presented at BERA Conference. University of Cambridge Local Examinations Syndicate. Available

online at <http://www.leeds.ac.uk/edocol/documents/000000061.htm>. Retrieved 5.9.2011.

Pearson, I. (1988). Tests as levers for change. In D. Chamberlain, & R. J. Baumgardner (eds.)

ESP in the Classroom. GB: Modern English Publications, 98-107.

Pollitt, A., Ahmed, A. & Crisp, V. (2007). The Demands of Examination Syllabuses and Question Papers. In P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (eds.) *Techniques for Monitoring the Comparability of Examination Standards*. Malta, 166-206.

Popham, W. J. (1983). Measurement as an Instructional Catalyst. In R. B. Ekstrom (ed.) *Measurement, Technology and Individuality in Education: Proceedings of the 1982 ETS Invitational Conference*. New Directions for Testing and Measurement, 17 San Francisco: Jossey-Bass, 19-30.

Popham, W. J. (1987). *The Merits of Measurement- Driven Instructions*. Phi Delta Kappa 68, 679-682.

Purpura, J. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press.

Rahman, T. (1996). *Language and Politics in Pakistan*. Karachi: Oxford University Press.

Rahman, T. (1997). The Medium of Instruction Controversy in Pakistan. *Journal of Multilingual and Multicultural Development*, 18(2), 145-154.

Rahman, T. (1999). *Language, Education and Culture*. Karachi: Oxford University Press.

Rahman, T. (2002). *Language, Ideology and Power: Language-Learning among the Muslims of Pakistan and North India*. Oxford: Oxford University Press.

Rahman, T. (2007). Language Policy, Multilingualism and Language Vitality in Pakistan. In A. Saxena and L. Borin (eds.) *Lesser-known Languages of South Asia: Status and Policies, Case Studies and Applications of Information Technology*. Berlin/ New York: Mouton de Gruyter.

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

OFSTED (1998). *Review of Secondary Education in England, 1993-1997*. London: OFSTED.

Shohamy, E. (1997). Second Language Assessment. In G. R. Tucker & D. Corson (ed.), *Second Language Education. Encyclopedia of Language and Education*, Vol. 4. Kluwer, Dordrecht.

Shohamy, E. (2000). *The Power of Tests: a Critical Perspective on the Uses and Consequences of Language Tests*. Harlow: Longman.

Soltero, S. W. (2004). *Dual Language: Teaching and Learning in Two Languages*. Boston: Pearson.

Skehan, P. (1991). Individual Differences in Second Language Learning. *Studies in Second Language Acquisition*, 13, 275-298.

Spolsky, B. (1976). *Language Testing: Art or Science?* Paper read at the 4th International Congress of Applied Linguistics. Stuttgart, Germany.

Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.

Sykes, A. O. (1992). An Introduction to Regression Analysis. Chicago Working Papers in

Law and Economics. Retrieved from

http://www.law.uchicago.edu/files/files/20.Sykes_Regression_0.pdf

Tattersall, K. (2007). A Brief History of Policies, Practices and Issues Relating to

Comparability. In P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (eds.)

Techniques for Monitoring the Comparability of Examination Standards. Malta, 43-91.

Taylor, L. (2004). Testing Times: Research Directions and Issues for Cambridge ESOL

Examinations. *TESOL Quarterly*, 38 (1) 141-146.

Taylor, L. (2005). Washback and Impact. *ELT Journal*, 59, 2.

Tymms, P., Coe, R. & Merrell, C. (2005). *Standards in English schools: Changes since 1997*

and the Impact of Government Policies and Initiatives. A Report for The Sunday Times,

London.

Van Dijk, T. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of*

Discourse. London: Longman.

Vernon, P. E. (1956). *The Measurement of Abilities* 2. London: University of London Press.

Wall, D. & Alderson, J. C. (1993). Examining Washback: the Sri Lankan Impact Study.

Language Testing, 10 (1) 41-69.

Wall, D. (2005). The Impact of High-Stakes Examinations on Classroom Teaching. In M.

Milanovic & C. J. Weir (eds.) *Studies in Language Testing Vol. 22* Cambridge:

Cambridge University Press.

Weigle, S.C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Weir, C. J. (1990). *Understanding and Developing Language Tests*. London: Prentice Hall.

Weir, C.J. (2003). A Survey of the History of the Certificate of Proficiency in English (CPE) in the Twentieth Century. In C.J. Weir and M. Milanovic (eds.) *Balancing Continuity and Innovation. A History of the CPE Examination 1913-2013, Studies in Language Testing, Vol 15*. CUP and Cambridge ESOL.

Weir, C. J. & Milanovic, M. (2003). *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*. UK: UCLES and Cambridge University Press.

Wiseman, S. (1961). *Examinations and English Education*. Manchester: Manchester University Press.

Wu, J. & Wu, R. (2010). Relating the GEPT Reading Comprehension Tests of CEFR. In M. Milanovic and C. J. Weir (eds.) *Aligning Tests with the CEFR*. Vol 33, Cambridge: UCLES/Cambridge University Press, 204-224.

APPENDICES

Appendix A

Matric English Syllabus (Reading & Writing)

Benchmark III: Write expository, persuasive, analytical essays, and personal narratives to produce a variety of academic and creative texts for various audiences.

Student Learning Outcomes

Grade IX & X

- Write and revise formal letters to people in extended social and academic environment for various purposes. Write the address on the envelope clearly and in proper **format**.
- Write and revise applications to people in extended environment using correct **format, layout and tons**.
- Analyze letters to editor to recognize effective ways of explaining and presenting an idea, a point, or an opinion.
- Write and revise letters to editor expressing Ideas, points, or opinions by supporting it with **anecdotes, quotations, examples, analogies, etc.**
- Analyze and compare various Informal and formal emails to note differences of conventions, **vocabulary, style and tone**.
- Write Informal and formal e mails in extended social and academic environment
- Analyze various forms currently required in extended social and academic environment.
- Fill in forms legibly, following instructions and supplying correct information.

Benchmark IV: Plan and draft their writing; revise and edit for various organization patterns of sequence, comparison, contrast, classification, cause and effect, logical flow of ideas through flexible and clear signal and reference words, **point of view**, supporting evidence, overall effect, appropriate punctuation and vocabulary.

Student Learning Outcomes

Grade IX & X

- Develop focus for own writing by identifying audience and purpose.
- Select and use a variety of pro strategies such as brainstorming, mind mapping, outlining etc
- Plan, draft and revise writing to ensure that It
 - Is focused, purposeful and reflects insight into the writing situation.
 - Has an organizational pattern that reflects a clear overall progression of Ideas through proper use of signal and reference words.
 - Uses writing strategies as are appropriate to the purpose of writing.
 - Has varied sentence structure and length.
 - Has a good command of language with precision of expression
- Proof read and edit their own, peers' and given texts for faulty sentence structure.
 - Faulty sentence structure.
 - Errors of subject / verb agreement.
 - Unclear pronoun reference.
 - Errors of correct word form.
 - Errors of punctuation and spelling.

Benchmark II: Write a variety of expository, persuasive, analytical essays, and personal narratives for different purposes and audiences.

**Student Learning Outcomes
Grade IX & X**

- Analyze to use in their own writing, features of an **expository** composition showing comparison and contrast between things, events, situations, places, action, idea or a problem:
 - Methods for organizing comparison and contrast.
 - Connectives of comparison and contrast.
 - Similes and **metaphors** for comparison and contrast.
- Write a personal **narrative** (autobiographical):
 - Include choice of time period in one's life.
 - Include clear statement of personality traits at that time.
 - Include significant details and an incident used to portray personality.
- Write a **persuasive** /argumentative essay on a given topic: -
 - Distinguish fact from opinion.
 - State an opinion on the topic.
 - List ideas and arguments that support opinion.
 - Organize ideas and supporting arguments in a clear, structured and logical manner.
 - Distinguish between language used for persuasion and propaganda.
 - Use **persuasive** language to enhance ideas.
 - Use special devices to support arguments (e.g. Appeal to logic through reasoning, appeal to emotion or ethical belief, relating a personal **anecdote** or **analogy**).
- Anticipate and respond to opposing arguments by defending **point of view** with factual evidence, quotations, expert opinion, logical reasoning, and commonly held beliefs.
- Analyze a simple sample book review of a story book to write one.
- Write a book review of a simple story book:
 - Write a summary of the book.
 - Gather information about the author.
 - Evaluate the plot, character and **setting**.
 - Prepare an outline for the book review.
 - Write, revise, and proof read the review.
- Analyze questions to write effective and focused answers of required length:
 - Mark **key words**.
 - Identify verbs and tenses.
 - Recognize **question types** such as
 - Literal / textual / factual
 - Interpretive
 - **Inferential**
 - Evaluative
 - Personal response
 - **Open ended**
- Use summary skills to write summary / **précis** of simple passages / poems.
- Use paraphrasing skills to paraphrase stanzas:
 - Mark **thought groups** in the stanza.
 - Restate the message in simple prose.
 - Replace poetic words with simple ones.

Competency 2: Writing Skills

Standard 1: All students will produce with developing fluency and accuracy, academic, transactional and creative writing, which is focused, purposeful and shows an insight into the writing process.

Benchmark 1: Analyze a variety of written discourse to use in their own compositions, techniques for effective text organization, development, and author's techniques that influence reader.

Student Learning Outcomes

Grade IX & X

- Develop focus for own writing by identifying audience and purpose.
- Select and use a variety of pre-writing strategies such as **brainstorming, mind mapping, outlining etc**
- Write a unified paragraph on a given topic to show
 - **Clear topic sentence** using specific words, vivid verbs, modifiers, etc.
 - Adequate supporting detail (example, illustration, definition, evidence, comparison, contrast, cause and effect, clarification, explanation) to develop the main idea.
 - Appropriate **pronoun-antecedent relationship** and **transitional devices** within a paragraph.
 - Chronological I sequential I spatial order of arranging detail.
 - Order of importance (most important to least important and vice versa, general to specific and vice versa).
- Analyze an essay to identify the general subject, main idea (a statement about the general subject), **key ideas, supporting details and transitional devices**.
- Write an essay on a general subject:
 - Write an introductory paragraph with a clear central thought.
 - Provide **key ideas** which prove, explain or support the central thought.
 - Use a separate paragraph for each key idea.
 - Incorporate evidence (facts, quotations, etc), examples (analogies, **anecdotes**, etc.), or different points of view (elaborating an idea/opinion) to support each key Idea.
 - Use appropriate **transitional devices** to connect ideas within and between paragraphs.
 - Add a closing or summary paragraph with a synthesis of central idea, syntheses of each supporting idea, a general concluding statement.
 - Use correct conventions of grammar and punctuation.
 - Use appropriate vocabulary.

Competency 1: Reading and Thinking Skills

Standard 2: All students will read and analyze literary text to seek information, ideas, enjoyment, and to relate their own experience to those of common humanity as depicted in literature.

Benchmark 1: Analyze short stories, poems, and essays; make connections between literary texts and their own lives.

Student Learning Outcomes

Grade IX & X

- Read text to make predictions about story line / content, characters, using **contextual clues** and prior knowledge.
- Analyze story elements: characters, events, **setting, plot, theme, tone, point of view**.
- Identify the speaker or **narrator** in a selection.
- Recognize the authors purpose and **point of view**.
- Read a text to
 - Make connections between their own lives and the characters, events, motives, and causes of conflict in texts.
 - Make **inferences** and draw conclusion about character using supportive evidence from the text.

- Compare characters in a literary selection to near similar one's in real life.
- Recognize **genres** of literature, fiction, nonfiction, poetry, **legend, myth**.
- Read a poem and give orally and in writing
 - Theme and its development.
 - Personal response with justification.
 - Paraphrase/ summary.
- Read and analyze how a writer/ poet uses language to
 - Appeal to the senses through use of **figurative language** including **similes** and **metaphors**.
 - Affect meaning through use of synonyms with different **connotations** and **denotations**.
 - **Create Imagery**.
- Read and recognize literary techniques such as repetition, personification and **alteration**.

Benchmark III: Analyze information from a visual cue or a graphic organizer to show complex processes, procedures, comparisons, contrasts, and cause and effect relationships

Student Learning Outcomes

Grade IX & X

- Interpret the situation in a **visual cue** and with the help of **mind map** giving vocabulary, structure and **sensory details** to write a short description.
- Analyze information in bar graphs, line graphs, and diagrams describing complex processes and procedures, comparisons and contrast to write a short written report.
- Organize information using various organizational patterns: sequence, comparison, contrast, classification, cause and effect.
- Recognize and use appropriate transitional words within and beyond paragraphs for better **coherence and cohesion**.
- Recognize and use appropriate conventions (**format, style, expression**)

Benchmark IV: Gather, analyze, evaluate and synthesize information to use for a variety of purposes including a research project using various aids and study skills.

Student Learning Outcomes

Grade IX & X

- Use dictionaries to
 - Locate **guide words**.
 - Locate **entry word**.
 - Choose appropriate word definition.
 - Identify pronunciation through pronunciation key.
 - Identify syllable division, and stress pattern.
 - Identify parts of speech.
 - Identify correct spellings.
 - Identify phrases through **key words**.
 - Recognize abbreviations used in a dictionary.
 - Locate phrases and Idioms.

- Comprehend notes on **usage**.
- Locate appropriate synonyms and antonyms in a thesaurus
- Use library skills to
 - Alphabetize book titles, words and names.
 - Locate fiction and non fiction books I books by subject.
 - Understand card catalogue.
 - Locate and use card catalogue.
 - Identify three kinds of catalogue cards I.e. author card, title card, subject card.
 - Use case and shelf labels in the library.
 - Use the reference section in the library.
 - Use computer catalogue.
- Utilize effective study strategies e.g. note-taking I note-making, writing a summary and creating a **mind map** to organize ideas.
- Use **textual aids** such as table of contents, footnote, glossary, preface, sub headings etc. to
 - Comprehend texts.
 - Identify and select relevant information in a book.
- Utilize appropriate informational sources including encyclopedias and internet sources

Benchmark II: Analyze complex processes, procedures, events, issues, and various viewpoints, applying reading comprehension and thinking strategies.

Student Learning Outcomes

Grade IX & X

- Use **pre-reading** strategies to
 - Predict the content of a text from topic / picture, title / headings, **key words** and **visuals** etc. by using prior knowledge, asking questions and **contextual clues**.
- Skim text to
 - Have general / idea of the text.
 - Infer theme / main idea.
- Apply **critical thinking** to interact with text and use intensive reading strategies (**while-reading**) to
 - **Scan** to answer short questions.
 - Make simple **inferences** using **context** of the text and prior knowledge.
 - Distinguish between what is clearly stated and what is implies.
 - Deduce meaning of difficult words from **context**.
 - Use **context** to infer missing words.
 - Read silently with comprehension and extract main idea and supporting detail.
 - **Scan** to locate an opinion.
 - Distinguish fact from opinion in letters to editors, texts supporting an opinion, etc.
 - Locate examples to support an opinion e.g a personal **anecdote**, quotations, examples and **analogy**, etc.
 - Recognize arguments and counter arguments.
 - Explore viewpoints / ideas and issues.
 - Follow instructions in maps or user instruction manuals and forms related to extended social and academic environment.

- Comprehend / interpret text by applying **critical thinking**.
- **Generate questions** to understand text.
- Use summary skills to
 - Extract salient points and develop a **mind map** to summarize a text.
 - Follow a process or procedure to summarize the information to transfer the written text to a diagram flow chart or **cloze** paragraph.
- Use critical thinking to respond orally and in writing to the text (**post-reading**) to
 - Give a personal opinion and justify stance related to viewpoints/ideas and issues in the text read.
 - Relate what is read to his or her own feelings and experiences.
 - Explore causes and consequences of a problem or an issue and propose various solutions.
 - Evaluate material read.
- Apply strategies to comprehend questions by marking **key words**, verbs and tenses in a variety of **question types**:
 - Literal / textual factual
 - Interpretive
 - **Inferential**
 - Evaluative
 - Personal response
 - **Open ended**
- Respond orally and in writing.

Student Learning Outcomes

Grades IX & X

Competency 1: Reading and Thinking Skills

Standard 1: All students will search for, discover and understand a variety of text types through tasks which require multiple reading and thinking strategies for comprehension, fluency and enjoyment.

Benchmark 1: Analyze patterns of text organization, and function of various devices used within and beyond a paragraph in a text.

Student Learning Outcomes

Grade IX & X

- Analyze passages in the text to identify the theme / general subject, key idea / central thought (a statement about the general subject), and **supporting details**.
- Analyze paragraphs to identify words, phrases or sentences that support the main idea through
 - Definition
 - Example / illustration
 - Cause and effect
 - Comparison and contrast
 - Facts, analogies, **anecdotes** and quotations.
- Identify and recognize the functions of
 - Pronoun – antecedent relationships.
 - Anaphoric and cataphoric references.
 - **Transitional devices** used for **coherence** and **cohesion** at **discourse level**.

➤ Analyze the order of arranging paragraphs:

- Chronological or spatial.
- General to specific, specific to general.
- Most important to least important and vice versa.

Appendix B:**O Level English Syllabus (Aims and Objectives)**

Cambridge O Level English Language
Syllabus code 1123
For examination in June and November 2013

3. Syllabus aims and objectives

3.1 Aims

A qualification in this syllabus demonstrates to universities and employers that candidates can communicate effectively in Standard English through:

1. **communicative competence:** the ability to communicate with clarity, relevance, accuracy and variety
2. **creativity:** the ability to use language, experience and imagination to respond to new situations, create original ideas and make a positive impact
3. **critical skills:** the ability to scan, filter and analyse different forms of information
4. **cross-cultural awareness:** the ability to engage with issues inside and outside own community, dealing with the familiar as well as the unfamiliar. (This is not an assessment objective but forms the context of writing tasks and reading passages.)

	Writing to:	Reading to:
Communicative competence	Communicate precisely and appropriately	Understand exact and implied meaning
Creativity	Develop ideas effectively	
Critical skills		Identify and respond to main ideas
Cross-cultural awareness	Reflect on the familiar	Have strategies to deal with the unfamiliar

Speaking and listening are not tested but the development of these vital communication skills is encouraged across the curriculum.

Reflecting the communication demands facing candidates in the real world, the syllabus distinguishes between **task** and **language** as the focus of Section 1 and Section 2 respectively in each paper:

Section	Focus	Writing	Reading
1	Task	Directed Writing	Reading for Ideas
2	Language	Creative Writing	Reading for Meaning

3. Syllabus aims and objectives

The Task aspect of Paper 1 is **Directed Writing**, where communication of key information is required to achieve a specific purpose for a certain audience in a particular situation. **Language** (as well as content) is tested in the **Creative Writing** section, where candidates have an opportunity to display their English language skills in order to express their opinion, experience or imagination.

The Task aspect of Paper 2 is **Reading for Ideas**, where, for example, scanning for and summarising specific information is required to achieve and convey a global understanding of a text. **Language** is tested in the **Reading for Meaning** section, where there is a greater demand for English language skills in order to demonstrate more in-depth understanding of a text.

In this way, it is hoped that candidates will develop strategies to be able to transfer these communication skills to other subjects and to their future careers/studies as they encounter a variety of texts and are required to make a positive impact through the written word.

3.2 Assessment Objectives

READING

- R1 Understand **explicit** meanings, through literal and vocabulary questions.
- R2 Understand **implicit** meanings and nuances of language, through inferential questions and questions on writer's craft.
- R3 **Scan and analyse** text, by identifying and summarising required information, such as similarities and differences, or advantages and disadvantages, or problems and solutions, or causes and effects, or actions and consequences.
- R4 Identify and respond to **main ideas** of a text, such as follow a sequence or argument, identify conclusion, distinguish fact from opinion, and give a personal response to a theme in a text.

WRITING

- W1 Communicate **appropriately**, with a clear awareness of purpose, audience and register.
- W2 Communicate **clearly** and develop ideas coherently, at word level, at sentence level and at whole text level.
- W3 Use **accurate** spelling, punctuation and grammar.
- W4 Communicate **creatively**, using a varied range of vocabulary, sentence structures and linguistic devices.

4 Description of components

4.1 Paper 1: Writing

1 hour 30 minutes, 60 marks

This paper has two sections and candidates answer on a separate answer sheet.

Section 1: Directed Writing (30 marks)

- Candidates are presented with a task, e.g. write a letter, speech, report, article, fit for purpose and relevant to the world of study, work or the community.
- Candidates should write 200–300 words to inform or persuade a particular audience.
- 15 marks are allocated for task fulfilment and 15 marks for language.

Section 2: Creative Writing (30 marks)

- This is an essay, testing language and content combined.
- Candidates answer one question from a choice of 5 narrative/descriptive/argumentative essay titles and should write 350–500 words.

Both sections test Assessment Objectives W1, W2, W3, W4.

4.2 Paper 2: Reading

1 hour 45 minutes, 50 marks

This paper has two sections and candidates answer on the question paper.

Section 1: Reading for Ideas (25 marks)

- Candidates scan a factual communication (or communications) of approximately 700 words – e.g. report(s), article(s), advertisement(s), email(s), letter(s).
- They identify and note down required information – e.g. similarities and differences, or causes and effects, or advantages and disadvantages, or problems and solutions, or actions and consequences. Only one example content point will be given as guidance to candidates.
- 15 marks are allocated for content points.
- Candidates use these notes to write a summary of 160 words. 5 marks are allocated for language.

This task tests Assessment Objective R3 (also implicitly R1, R2).

5 Marking band descriptors for Paper 1

Band 5 (16–14 marks)

- Good understanding of purpose.
- Clear awareness of situation and audience.
- Format entirely appropriate.
- All required points developed in detail, fully amplified and well organised.
- Given information well used to justify personal opinion and interpretation.
- Tone and register entirely appropriate.

Band 2 (12–10 marks)

- An understanding of purpose.
- An awareness of situation and audience.
- Format appropriate.
- All required points addressed but not always developed in detail.
- Given information organised to support personal opinion.
- Tone and register appropriate.

Band 3 (9–7 marks)

- Some understanding of purpose.
- Some awareness of situation and audience.
- Format generally appropriate.
- At least two required points addressed (and partially/fully developed).
- Given information may not be logically used to support opinion.
- Tone usually appropriate, although there may be slips of register.

Band 4 (6–4 marks)

- Only partial understanding of purpose.
- Some confusion as to situation and audience.
- Format may be inappropriate.
- At least one of the required points addressed (and partially/fully developed).
- Given information may be used irrelevantly.
- Tone may be uneven.

5. Marking band descriptors for Paper 1

Band 1 (5–4 marks)

- Highly accurate, apart from very occasional slips.
- Sentence structures varied for particular effects.
- Verb forms largely correct and appropriate tenses consistently used.
- Vocabulary wide and precise.
- Punctuation accurate and helpful.
- Spelling accurate, apart from very occasional slips.
- Paragraphs have unity, are linked, and show evidence of planning.

Band 2 (13–12 marks)

- Accurate; occasional errors are either slips or caused by ambition.
- Sentence structures show some variation to create some natural fluency.
- Occasional slips in verb forms or tense formation, but sequence consistent and clear throughout.
- Vocabulary precise enough to convey intended shades of meaning.
- Punctuation accurate and generally helpful.
- Spelling nearly always accurate.
- Paragraphs have unity, are usually linked and show some evidence of planning.

Band 3 (11–10 marks)

- Mostly accurate; errors from ambition do not mar clarity of communication.
- Some variety of sentence structures, but tendency to repeat sentence types may produce monotonous effect.
- Errors may occur in irregular verb forms, but control of tense sequence sufficient to sustain clear progression of events or ideas.
- Simple vocabulary mainly correct; errors may occur with more ambitious words.
- Spelling of simple vocabulary accurate; some errors in more ambitious words.
- Punctuation generally accurate and sentence separation correctly marked, but errors may occur e.g. with direct speech.
- Paragraphs may show some unity, although links may be absent or inappropriate.

Band 4 (9–8 marks)

- Sufficiently accurate to communicate meaning, with patches of clear, accurate language.
- Some variety of sentence length and structure, not always for particular purpose.
- Errors in verb forms and tense consistency may cause uncertainty in sequence of events or disturb ease of communication.
- Vocabulary usually adequate to convey intended meaning; idiom may be uncertain.
- Spelling of simple vocabulary accurate, errors in more difficult words.
- Punctuation used but not always helpful; occasional sentence separation errors.
- Paragraphs used but may lack unity or coherence.

5. Marking band descriptors for Paper 1

Band 5 (7–6 marks)

- Overall meaning never in doubt, but errors sufficiently frequent and serious to hamper precision and distract reader from content.
- Some simple structures accurate but script unlikely to sustain accuracy for long.
- Errors in verb forms and tenses will sometimes confuse sequence of events.
- Vocabulary limited, either too simple or imperfectly understood; some idiomatic errors likely.
- Spelling of simple vocabulary accurate, frequent errors in more difficult words.
- Simple punctuation usually accurate, but there may be frequent sentence separation errors.
- Paragraphs used haphazardly.

Band 6 (5–4 marks)

- Many serious errors of various kinds of 'single-word' type (i.e. they could be corrected without re-writing the sentence); communication established, although weight of error may cause some 'blurring'.
- Sentences probably simple and repetitive in structure.
- Frequent errors in verb forms and haphazard changes of tense confuse meaning.
- Vocabulary conveys meaning but likely to be simple and imprecise; significant idiomatic errors.
- Spelling may be inconsistent.
- Punctuation and paragraphing may be haphazard or non-existent.

Band 7 (3–2 marks)

- Sense usually decipherable but some error will be 'multiple' (i.e. requiring the reader to re-read and re-organise); meaning may be partly hidden by density of linguistic error.
- Unlikely to be more than a few accurate sentences, however simple, in the whole composition.

Band 8 (1–0 mark)

- Scripts almost entirely or entirely impossible to recognise as pieces of English writing; whole sections make no sense at all.
- Where occasional patches of relative clarity are evident, 1 mark should be given.
- The mark of 0 is reserved for scripts that make no sense at all from beginning to end.

5 Marking band descriptors for Paper 1

SECTION 2 MARK

Band 1 (30–27 marks)

- **Highly accurate**, apart from very occasional slips.
- **Sentence structure** varied for particular effects.
- **Verb forms** largely correct and appropriate tenses **consistently used**.
- **Vocabulary** wide and precise.
- **Punctuation** accurate and helpful.
- **Spelling** accurate apart from very occasional slips.
- **Paragraphs** have unity, are linked, and show evidence of planning.

Consistently relevant. Interest aroused and sustained. Tone and register entirely appropriate.

Discursive essays are well developed, logical, even complex, in argument.

Descriptive essays have well-developed images helping to create complex atmospheres.

Narratives are complex, sophisticated, possibly tense, and may contain devices such as flashbacks.

Band 2 (26–23 marks)

- **Accurate: occasional errors** are either slips or caused by ambition.
- **Sentence structures** show some variation to create some natural fluency.
- Occasional slips in verb forms or tense formation but sequence consistent and clear throughout.
- **Vocabulary** wide and precise enough to convey intended shades of meaning.
- **Punctuation** accurate and generally helpful.
- **Spelling** nearly always accurate.
- **Paragraphs** have unity, are usually linked and show some evidence of planning.

Relevant. Interest aroused and mostly sustained. Tone and register appropriate.

Discursive essays have clearly-defined, cohesive, logical stages in their argument.

Descriptive essays have interesting images and range of detail, helping to create effective atmospheres.

Narratives have effective detail creating character or setting, and may contain some sense of climax.

Band 3 (22–19 marks)

- **Mostly accurate**; errors from ambition do not mar clarity of communication.
- **Some variety of sentence structures**, but tendency to repeat sentence types may produce monotonous effect.
- Errors may occur in irregular verb forms, but **control of tense sequence** sufficient to sustain clear progression of events or ideas.
- **Simple vocabulary** mainly correct; errors may occur with more ambitious words.
- **Punctuation** generally accurate and **sentence separation** correctly marked, but errors may occur e.g. with direct speech.
- **Spelling of simple vocabulary** accurate; some errors in more ambitious words.
- **Paragraphs** may show some unity, although links may be absent or inappropriate.

5. Marking band descriptors for Paper 1

Relevant. Some interest aroused, although there may be some lack of originality and/or planning. **Tone usually appropriate**, although there may be slips of register.

Discursive essays make a series of relevant points, with some being developed; linking of ideas may be insecure.

Descriptive essays have satisfactory images, ideas and details which help to create atmosphere. **Narratives** are straightforward with proper sequencing of sentences

Band 4 (18–15 marks)

- **Sufficiently accurate to communicate meaning**, with patches of clear, accurate language.
- **Some variety of sentence length and structure**, not always for particular purpose.
- **Errors in verb forms and tense consistency** may cause uncertainty in sequence of events or disturb ease of communication.
- **Vocabulary usually adequate to convey intended meaning**; idiom may be uncertain.
- **Punctuation used but not always helpful**; occasional sentence separation errors.
- **Spelling of simple vocabulary accurate**; errors in more difficult words.
- **Paragraphs used but may lack unity or coherence**.

Attempt to address topic but there may be digressions or failures of logic. May lack liveliness and interest.

Tone may be uneven.

Discursive essays have mainly relevant points but may be only partially developed, with some repetition.

Descriptive essays have some detail but may rely too much on narrative.

Narratives are largely a series of events with only occasional details of character and setting.

Band 5 (14–11)

- **Overall meaning never in doubt**, but errors sufficiently frequent and serious to hamper precision and distract reader from content.
- **Some simple sentence structures accurate but script unlikely to sustain accuracy for long**.
- **Errors in verb forms and tenses** will sometimes confuse sequence of events.
- **Vocabulary limited**, either too simple or imperfectly understood; some idiomatic errors likely.
- **Simple punctuation usually accurate**, but there may be frequent sentence separation errors.
- **Spelling of simple vocabulary accurate**, frequent errors in more difficult words.

Paragraphs used haphazardly.

Some relevance. Some interest.

Tone may be inconsistent.

Discursive essays make a few points but development is simple and not always logical; some obvious repetition of ideas.

Descriptive essays are relevant but lack scope or variety. **Narratives** are simple, everyday or immature.

5.M Descriptors for Paper 1

a
r
k

Band 6 (10–7)

- Many serious errors of various kinds of 'single-word' type (i.e. they could be corrected without rewriting the sentence); communication established, although weight of error may cause some 'blurring'.
- Sentences probably simple and repetitive in structure.
- Frequent errors in verb forms and haphazard changes of tense confuse meaning.
- Vocabulary conveys meaning but likely to be simple and imprecise; significant idiomatic errors.
- Punctuation and paragraphing may be haphazard or non-existent.
- Spelling may be inconsistent.

b little relevance. A little interest.

Some recognition of appropriate tone.

In Discursive essays only a few points are discernable and the argument progresses only here and there.

In Descriptive essays the overall picture is unclear.

Narratives are very simple and may narrate events indiscriminately.

Band 7 (6–3)

D Sense usually decipherable but some error will be 'multiple' (i.e. requiring the reader to re-read and re-organise); meaning may be partly hidden by density of linguistic error.

- Unlikely to be more than a few accurate sentences, however simple, in the whole composition.
- Little relevance or interest. Tone may be inappropriate.

D In Discursive essays only a very few points are discernable and the argument barely progresses. In Descriptive essays the overall picture is very unclear.

Narratives are extremely simple and may narrate events indiscriminately.

Band 8 (2–0)

S Scripts almost entirely or entirely impossible to recognise as pieces of English writing; whole sections make no sense at all.

- Where occasional patches of relative clarity are evident, 2 or 1 mark(s) should be given.
- The mark of 0 is reserved for scripts that make no sense at all from beginning to end.

C Discursive essays are rarely relevant and may well be disordered, as are Descriptive essays and Narratives.

i
p
D

Appendix C: O Level Paper 1: Writing

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

ENGLISH LANGUAGE

Paper 1 Writing

1123/11

October/November 2011

1 hour 30 minutes

Additional Materials: Answer Booklet/Paper

**READ THESE INSTRUCTIONS FIRST**

If you have been given an Answer Booklet, follow the instructions on the front cover of the Booklet. Write your Centre number, candidate number and name on all the work you hand in.

Write in dark blue or black pen.

Do not use staples, paper clips, highlighters, glue or correction fluid.

Answer both **Section One** and **Section Two**.

At the end of the examination, fasten all your work securely together.

This document consists of 3 printed pages and 1 blank page.



UNIVERSITY of CAMBRIDGE

Section 1: Directed Writing

You are advised to write between 200 and 300 words. Total marks for this part: 30.

Task

Recently you witnessed a major disturbance at a railway station when many people were injured. You are asked by the local police to write an **account** of what you saw.

Write your **account**. You must include the following:

1. when and where the incident happened **and** how close you were to the scene
2. the cause of the disturbance **and** what exactly happened
3. some of the ways in which people tried to help.

Cover all three points above in **detail**. You should make your account informative and helpful for the police to show you are a reliable witness. Start your account 'To the Police...' and remember to add your signature and the date.

Section 2: Creative Writing

Begin your answer on a fresh page.

Write on **one** of the following topics.

At the head of your essay put the number of the topic you have chosen.

You are advised to write between 350 and 500 words. Total marks for this part: 30.

- 1 Describe a time when you prefer to be alone and a time when you like to be part of a crowd. (Remember that you are describing the atmosphere and your feelings, not telling a story.)
- 2 What are the advantages and disadvantages of attending either a small school with few students or a large school with many students?
- 3 Write a story which includes the sentence: 'As we entered the building, the other people smiled as if they knew something we did not.'
- 4 Bullies.
- 5 Write a story about someone who returned to a village or town after a long time away. (You should include full details of why the person went away to show that it is an important part of your story.)

BLANK PAGE

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

University of Cambridge International Examinations is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

Appendix C: O Level Paper 2: Reading

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

ENGLISH LANGUAGE
Paper 2 Reading
INSERT

1123/21

October/November 2011

1 hour 45 minutes

READ THESE INSTRUCTIONS FIRST

This insert contains the two reading passages.

Passage 1 – Mobile Phones

4. One of the most obvious results of our exciting technological age has been the invention of the mobile phone. Many people – even children as young as seven or eight – have one. But are mobile phones necessarily beneficial to society? 5

5. Mobiles can be seen as a way for young people to keep in touch with their parents if they are out with friends and, as newspapers make increasing reports of violence, particularly in inner city areas, this must be an advantage. If students are on a school outing and have to be met by parents on their return to school, they can phone at short notice from, say, the bus to give details of exact timings. Mobiles encourage friendships and some students speak to their friends several times a day. Parents may be happy that their children are not using the house telephone and incurring charges which they, as parents, have to meet – an issue which, in the past, was often a source of family conflict. Most young people have 'pay as you go' mobiles, and this encourages budgeting, surely a useful skill to have in later life. 10

6. Because a mobile is the personal property of its owner, direct contact can be made, and there is less risk of disturbing an entire household engaged in some family activity, such as having dinner. Another obvious advantage is that people can be contacted wherever they are, clearly 15 a benefit if a friend or spouse wishes to communicate that he is stuck in traffic, or that the train is running late. In situations where it is impossible for either the caller or the recipient to speak, mobiles can be used to send short text messages, which are both convenient and relatively cheap. Many mobiles are also able to take photographs; this multi-tasking means that there is no need to purchase a camera, and thus a saving is made. Sophisticated mobiles 20 even allow their owners to send and receive e-mails, thus providing an alternative means of communication. Small businesses flourish because of mobile phones. A plumber, for example, can receive bookings for future work while he is working on another job; there is no need to sit at home waiting for phone calls or to employ a secretary in an office.

4. However, there are downsides to mobile phones. People are discouraged from planning 25 ahead. It is a rather pathetic sight to see a wife phoning from the supermarket to ask her husband what he would like for dinner. People might also stop making simple decisions for themselves with, say, a teenager phoning his mother to ask if he may take a drink out of the fridge. Then there is the general disturbance which can be caused: there is nothing worse than people on public transport conducting loud and animated conversations on their mobiles with 30 scant regard for their fellow passengers. Some countries, for example France, have tackled this problem by insisting that mobiles are switched off on trains; all countries should follow this example. Unfortunately, owning a mobile phone prevents some people from concentrating on the present moment. Picture the scene: a young couple are dining in a fine restaurant but, instead of being absorbed in each other's company, each of them is speaking to someone 35 else via a mobile phone.

5. In schools, there can be regrettable competition among students to have the most up-to-date mobile. Some students are distracted into texting friends in class, which clearly hinders their education. In cases where mobiles are confiscated because of such misuse, there are 40 security issues for teachers – who is responsible for this expensive equipment once it has been confiscated? And, of course, teachers complain vociferously about text message spelling and grammar which, they say, encourage sloppy use of language: 'CU' is enough to raise the blood pressure of an English teacher to dangerous levels. Furthermore, they complain that mobile phones, along with e-mails, have destroyed the art of letter writing. Parents complain that their children contact their friends – with whom they have been all day in school – several 45 times during the evening via text or call instead of doing their homework. 'What do they have to talk about?' is the perceived mystery.

6. Is the mobile phone a blessing or a curse? Perhaps there is no clear cut answer to this and, as with many issues, moderation is essential.

Passage 2 – An Otter in the Air

(The writer describes his experience of taking an animal with him on a plane in the days when this was permitted.)

1 The plane was waiting to take off; as I rushed through the airport, carrying the box which was Mij's temporary home, my mind boggled at the thought of the next few hours. I was trying to hold down the lid of the box with one hand and, with the other, to force back the screw into the splintered wood. 5

2 The other passengers stared at me inquisitively as I struggled onto the plane with my horrifically vocal box. I was anxious to see who would be my immediate neighbour and was dismayed to find an elegantly dressed middle-aged woman. Such a person, I thought, would have little tolerance and certainly no sympathy for the scruffy otter cub that would so soon be her travelling companion. For the moment the lid held, and there was a brief silence from within the box. 10

3 Worried about how I would keep Mij under control throughout the flight, I had brought a bottle of water and a parcel of fish, and with these scant resources I prepared to withstand a siege. Aware of the fact that I could not keep Mij's presence a secret for long, and of the need to keep the fish in a cool place, I spoke to the stewardess; I daresay I was not too coherent but she took it all in her graceful stride and received the mundane parcel of fish as though I were travelling 15 royalty depositing a jewel case into her safe keeping. When the stewardess suggested I remove my pet from its box and have it on my knee, my neighbour, surprisingly, had no objection. 15

5. For the first hour or so Mij slept in my lap. However, otters are extremely bad at doing nothing. There is, I am convinced, something positively provoking to an otter about order in any form and, the greater the state of confusion they can create, the more contented they feel. A room 20 is not properly habitable until they have turned everything in it upside down. One of these moods descended on Mij. It began comparatively innocuously when he jumped down from my lap – inertia was not for him. Then he turned his attention to the box, which was on the floor, filled with wood shavings for him to sleep on. He put his head into the box and began to throw the wood shavings out backwards at enormous speed; then he got in bodily and lay 25 on his back, using all four feet in a pedalling motion to hoist out the remainder. With his teeth, he yanked back the zip on my neighbour's bag and was in head first, throwing out all the personal paraphernalia of air travel. But there was worse to come. 25

5 With a rebellious and eel-like wriggle, Mij disappeared at high speed. I could follow his progress by the wave of disturbance down the passageway. There were squawks and shrieks; then a 30 woman halfway down the plane stood up on her seat screaming, 'A rat! A rat!' The stewardess reached her and, within a matter of seconds, the woman was seated again, smiling benignly. That goddess, I believe, could have controlled a panic-stricken crowd single-handedly. 30

6 By now I was in the passageway myself and, catching sight of Mij's tail disappearing beneath the legs of a portly gentleman, I tried a flying tackle, landing flat on my face. I missed Mij's tail, but 35 found myself grasping the foot of the gentleman's female companion. The gentleman gave me a long silent stare. I staggered up, babbling my apology; the man was so utterly expressionless that even in my hypersensitive mood I could deduce no meaning from him whatsoever. 35

7 The stewardess came to my assistance once again. 'Perhaps,' she said with her charming smile, 'you should return to your seat and I will bring the animal to you.' I heard the ripple of flight and pursuit passing up and down the body of the plane, but I could see little. I was craning my neck trying to follow the hunt when suddenly I heard from my feet a distressed chitter of recognition and welcome. In all the strange world of the aircraft I was the only familiar thing to be found, and in that spontaneous return was sown the seed of the absolute trust that Mij accorded me for the rest of his life. 40 45

BLANK PAGE

Appendix C: O Level Paper 2: Reading

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary LevelCANDIDATE
NAME

--	--	--	--	--

CENTRE
NUMBER

--	--	--	--	--

CANDIDATE
NUMBER

--	--	--	--

ENGLISH LANGUAGE

1123/21

Paper 2 Reading

October/November 2011

1 hour 45 minutes

Candidates answer on the Question Paper.

Additional Materials: Insert

READ THESE INSTRUCTIONS FIRST

Write your Centre number, candidate number and name on all the work you hand in.

Write in dark blue or black pen.

Do not use staples, paper clips, highlighters, glue or correction fluid.

DO NOT WRITE IN ANY BARCODES.

Answer all questions in both Section 1 and Section 2.

The insert contains the two reading passages.

Mistakes in spelling, punctuation and grammar may be penalised in any part of the Paper.

Dictionaries are not permitted in this examination.

At the end of the examination, fasten all your work securely together.

The number of marks is given in brackets [] at the beginning or end of each question or part question.

For Examiner's Use	

This document consists of 7 printed pages, 1 blank page and 1 insert.

Section 1: Reading for Ideas

For
Examiner's
Use

Read Passage 1 in the insert and answer **all** the questions below in the order set.

1 (a) Notes [15 marks]

Identify and write down the advantages and disadvantages of mobile phones, as described in the passage.

USE ONLY THE MATERIAL FROM PARAGRAPH 2 TO PARAGRAPH 5 INCLUSIVE.

At this stage, you need NOT use your own words. To help you get started, the first point in each section of notes is done for you. You will be awarded up to 15 marks for content points.

MAIN POINTS
Advantages
<ul style="list-style-type: none">• <i>Young people can keep in touch with their parents</i>
Disadvantages
<ul style="list-style-type: none">• <i>People don't plan ahead</i>

(b) Summary [5 marks]

Now use your notes to write a summary of the advantages and disadvantages of mobile phones, as described in the passage.

For
Examiner's
Use

This time, you will be awarded up to 5 marks for using your own words wherever possible and for accurate use of language.

Your summary, which must be in continuous writing (not note form), must be no longer than 160 words, including the 10 words given below. Begin your summary as follows:

It is clear that mobile phones are useful devices because

No. of words

2 From paragraph 4, select and write down **two** opinions.

One opinion is [1]

Another opinion is [1]

3 From the whole passage, which of the following statements is correct? Tick the box you have chosen.

The writer is totally against mobile phones.

The writer thinks mobile phones can be useful.

The writer is in favour of mobile phones.

[1]

4 From your **own** knowledge or experience, suggest **two** advantages of mobile phones **OR** two disadvantages, **OR** **one** advantage and **one** disadvantage. Do not refer to specific examples from the passage in your answer.

One advantage/disadvantage is [1]

One advantage/disadvantage is [1]

Total for Section 1 [25]

Section 2: Reading for Meaning

For
Examiner's
Use

Read Passage 2 in the insert and answer all the questions below in the order set.

From paragraph 1

5 (a) Why did the writer rush through the airport?

..... [1]

(b) Why do you think the writer was trying 'to hold down the lid of the box'?

..... [1]

From paragraph 2

6 (a) Give two reasons why, according to the writer, Mij would not be considered an attractive travelling companion.

(i)

(ii)

[2]

(b) What, according to the writer, would be the woman's attitude to Mij? Answer in your own words.

.....
.....
.....

[2]

From paragraph 3

7 (a) The writer 'spoke to the stewardess'. Explain what he told her and what he asked her to do.

.....
.....
.....

[2]

(b) Without using the words of the passage, explain fully what was unusual about the way the stewardess reacted to the writer's request.

.....
.....
.....

[2]

From paragraph 4

8 (a) 'Otters are extremely bad at doing nothing'. Pick out and write down the single word used later in the paragraph which continues the idea of 'doing nothing'.

..... [1]

(b) 'One of these moods descended on Mij'. What kind of 'mood' was this?

..... [1]

(c) Why did Mij climb into the box?

..... [1]

From paragraph 5

9 (a) There was a 'wave of disturbance' down the passageway. Explain fully what was happening here.

.....
.....
..... [2]

(b) Why does the writer describe the stewardess as a 'goddess'?

.....
..... [1]

From paragraph 6

10 Give two reasons why the writer 'could deduce no meaning' from the gentleman who stared at him.

(i)

(ii)

[2]

From paragraph 7

11 (a) How was the writer eventually re-united with Mij?

..... [1]

(b) How did his re-union with Mij affect the writer's relationship with him in the future?

..... [1]

From the whole passage

12 Choose five of the following words. For each of them give one word or short phrase (of not more than seven words) which has the same meaning that the word has in the passage.

(a) inquisitively (line 5)

(e) order (line 19)

(b) scant (line 12)

(f) hoist (line 26)

(c) coherent (line 14)

(g) paraphernalia (line 28)

(d) provoking (line 19)

(h) craning (line 42)

Word chosen	Answer	
-------------	--------	--

() [1]

() [1]

() [1]

() [1]

() [1]

[5]

Total for Section 2 [25]

Appendix D**Matric English Papers**

Roll Number _____

Answer Sheet No. _____

Sig of Candidate _____

Sig of Invigilator _____

ENGLISH COMPULSORY SSC-I
SECTION – A (Marks 15)

Time allowed: 20 Minutes

NOTE: Section-A is compulsory. All parts of this section are to be answered on the question paper itself. It should be completed in the first 20 minutes and handed over to the Centre Superintendent. Deleting / overwriting is not allowed. Do not use lead pencil.

Q. 1 Circle the correct option i.e A / B / C / D. Each part carries one mark.

(i) Who was Hazrat Salman Farsi (RA)?
 A. A Caliph B. A Commander C. A Slave
 D. A Saint

(ii) Where should animals be branded according to Hazrat Muhammad (PBUH)?
 A. Face B. Back C. Legs
 D. Head

(iii) When does each of the Chinese month start?
 A. 1st of every month B. 1st of January
 C. New Season D. On the day of the new moon

(iv) _____ were the earliest people who introduced the marathon race and many other sports.
 A. Greeks B. Japanese C. Chinese D. Americans

(v) Abou bin Adhem requested the angel to write his name in the list of those who _____
 A. Love the Lord B. Love His fellow men
 C. Work hard D. Love themselves

(vi) Traffic rules apply _____ everybody who is on the road.
 A. On B. Over C. To D. In

(vii) Choose the correct spelling:
 A. Equavalent B. Aquivalent C. Equivalent D. Equivalent

(viii) Which of the following words is NOT synonym to the word 'Opponent'?
 A. Friend B. Enemy C. Foe D. Against

(ix) Knowledge strengthens our belief. The underlined word is a / an _____
 A. Noun B. Verb C. Adjective D. Adverb

(x) Insects that live and work together are called _____ insects.
 A. Small B. Bees C. Social D. Drones

(xi) 'He is the young man who saved my life'. The underlined word is a / _____
 A. Personal Pronoun B. Interrogative Pronoun
 C. Indefinite Pronoun D. Relative Pronoun

(xii) 'He was reading clearly'. The underlined word is an _____
 A. Adverb of Manner B. Adverb of Place
 C. Adverb of Time D. None of these

(xiii) Which of the following is a **CORRECT** sentence?
 A. I am surprised on your behavior. B. I am surprised at your behavior.
 C. I am surprised over your behavior. D. I am surprised in your behavior.

(xiv) Which of the following is **NOT** a sentence?
 A. Why are such rumours spread? B. The story of the accident was being told.
 C. You did not fulfil. D. I did not make fun of him.

(xv) 'Our team has won the match'. What tense is it?

A. Present Indefinite
C. Past Indefinite

B. Present Perfect
D. Past Perfect

For Examiner's use only:

Total Marks: 15

Marks Obtained: _____

ENGLISH COMPULSORY SSC-I

Time allowed: 2:40 Hours

Total Marks Sections B and C : 60

NOTE: Answer all the questions from Sections 'B' and 'C' on the separately provided answer book. Use supplementary answer sheet i.e Sheet-B if required. Write your answers neatly and legibly.

SECTION – B (Marks 39)

**Q. 2 Answer any SIX of the following parts in about 30 to 40 words each. Each part carries 3 marks
(6 x 3 = 18)**

- (i) Write an instance that you have read in the lesson, "Kindness to Living Things" about kindness to animals from the life of Hazrat Muhammad (PBUH).
- (ii) Why did the Chinese King send his son to the great master Pan Ku?
- (iii) What game was used for utilitarian purpose in the past?
- (iv) Describe the process by which bees gather food.
- (v) How do the roots of plants get water from the soil? Explain in the light of the lesson. 'What happens to the Rain?'
- (vi) How did Hina get sick in the lesson. 'Doctor's Advice'?
- (vii) How does the poet feel in the company of daffodils? (Refer to the poem. 'Daffodils')
- (viii) What are the causes of traffic accidents?

Q. 3 A. Paraphrase any ONE of the following stanzas:

(03)

- (i) Yet, of the dark I have no fear, (ii) An angel writing in a book of gold,
But feel as safe as when it's light, Exceeding peace had made bin
For I know God is with me there, Adhem bold, And to the presence
An He will guard me through the in the room he said,
night. "What writest thou?"

B. Read the following stanza carefully and answer the questions given at the end: (06)

But Rebecca recollected,
She was taught deceit to shun,
And the moment she reflected.
Told her mother what was done.

Questions :

- (i) Which poem has the above stanza been taken from? Also give the name of the poet.
- (ii) Explain the phrase. 'deceit to shun'.
- (iii) What did Rebecca decide to do in the end?

Q. 4 Change the voice of any FOUR of the following sentences:

(04)

- (i) Who did this?
- (ii) Grammar is taught to us by Mr. Ahmad.
- (iii) He was taken to the hospital by his friends.
- (iv) The enemy has defeated our army.
- (v) I shall order the carriage.
- (vi) I am watching you very carefully.
- (vii) The exhibition was opened by the Governor.

Q. 5 Use any FOUR of the following phrasal verbs in sentences:

(04)

- (i) Carry on (ii) Knock at (iii) See through (iv) Talk over

(v) Take to (vi) Get off (vii) Break down

Q. 6 Punctuate and capitalize the following para / line (s): (04)

We have had women politicians as prime minister education minister and also as heads of various institutions of the education department

SECTION – C (Marks 21)

Q. 7 Write a letter to your mother telling her about your health. (08)

-OR-

Write a letter to the editor of a newspaper telling about the drug addicts in your area.

Q. 8 Write a paragraph of about 50 to 70 words on “Recent Flood in Pakistan” with the help of the following word bank: (08)

Heavy rains – snow melted – cause of flood – a lot of destruction in the Northern areas – people homeless – without shelter, food – now diseases.

Q. 9 Translate any FIVE of the following Urdu sentences into English (05)

(iii)	(ii)	(i)
(vi)	(v)	(iv)
		(vii)

-OR-

Write a dialogue between two friends on the cricket match Pakistani team has won.

Roll Number _____

Answer Sheet No. _____

Sig of Candidate _____

Sig of Invigilator _____

ENGLISH COMPULSORY SSC-II

SECTION – A (Marks 15)

Time allowed: 20 Minutes

NOTE: Section-A is compulsory. All parts of this section are to be answered on the question paper itself. It should be completed in the first 20 minutes and handed over to the Centre Superintendent. Deleting / overwriting is not allowed. Do not use lead pencil.

Q. 1 Circle the correct option i.e A / B / C / D. Each part carries one mark.

(i) The eight lunar month is _____
A. Ramadan B. Rajab C. Shaban D. Safar

(ii) Synonym of 'Freedom' is _____
A. Independence B. Dependence C. Slavery
D. Dependent

(iii) Hazrat Ayesha (RA) remained with the Holy Prophet (PBUH) for _____
A. Eight years B. Ten years C. Twelve years
D. Thirteen years

(iv) The poem 'A Nation's Strength' has been composed by _____
A. Cecil Alexander B. Henry Wotton
C. R.W Emerson D. William Wordsworth

(v) Which is the opposite of 'Generous'?
A. Ungrateful B. Gentle C. Miser D. Famous

(vi) You are writing a letter to your friend. How would you salute him?
A. Dear Friend B. Dear Madam
C. Respected Friend D. All of these

(vii) 'It is a fault that will right itself'. The underlined word is a / an _____
A. Verb B. Noun C. Adjective D. Adverb

(viii) All the intoxicants have _____ damaging for health.
A. Proof B. Proven C. To prove D. None of these

(ix) What does the word 'Potent' mean?
A. Spoil B. Glorify C. Strong D. Lose

(x) 'He does not take pride in his piety'. Which tense is it?
A. Present Perfect B. Present Indefinite
C. Present Continuous D. Present Perfect Continuous

(xi) Which of the following is the correct sentence?
A. One of the boy had been absent. B. One of the boys were absent.
C. One of the boy was absent. D. One of the boys was absent.

(xii) Which of the following was the meeting place of Muslims and the headquarter of the administration?
A. Ka'aba B. Arafat C. Masjid-e-Nabvi
D. Masjid-ul-Haram

(xiii) Which of the following is a Relative Pronoun?
A. These B. Myself C. Them D. Anyone

(xiv) Crafts were the main source of _____ for the people of the country.
A. Cooperation B. Fame C. Leisure D. Livelihood

(xv) 'He did this Herculean task himself and proved his worth'. This is a _____
A. Complex sentence B. Simple sentence
C. Compound sentence D. Complex compound sentence

For Examiner's use only:

Total Marks: 15

Marks Obtained: _____

ENGLISH COMPULSORY SSC-II

Time allowed: 2:40 Hours

Total Marks Sections B and C : 60

NOTE: Answer all the questions from Sections 'B' and 'C' on the separately provided answer book. Use supplementary answer sheet i.e Sheet-B if required. Write your answers neatly and legibly.

SECTION – B (Marks 34)

Q. 2 Answer any **FIVE** of the following parts in about 30 to 40 words each. Each part carries three marks ($5 \times 3 = 15$)

- (i) Hazrat Muhammad (PBUH) is the greatest educationalist of the world. Discuss with reference to the lesson, 'Prophet Muhammad (PBUH) as an Educationist.'
- (ii) What is the importance of a Lok Mela?
- (iii) Mention any three clauses of the treaty of Medina.
- (iv) Can computers take over the world in future?
- (v) How can we do away with the drugs from the society?
- (vi) Why did the Khalifa invite certain people to dinner?
- (vii) What is required by a person who wishes to lead a 'well balanced life'?

Q. 3 A. Paraphrase any **ONE** of the following stanzas:

(03)

- (i) Then deem it not an idle thing,
A pleasant word to speak.
The face you wear, the thoughts you bring,
A heart may heal or break.
- (ii) Who ran to help me when I fell,
And would some pretty story tell,
Or kiss my head to make me well?
My Mother

B. Read the following stanza carefully and answer the questions given at the end:

(06)

Whose passions not his masters are,
Whose soul is still prepared for death,
United unto the world by care,
Of public fame or private breath.

Questions :

- (i) What qualities are required for a happy life according to this stanza?
- (ii) Is the soul of a happy man prepared for death? Explain.
- (iii) Explain the words 'Public Fame' and 'Private Breath'?

Q. 4 Read the following passage and answer the questions given at the end:

The age of steam has yielded place to the age of electricity. We have harnessed and introduced it into the service of man. The electric telegraph bears our message to and brings us news from all quarters of the world. Submarine cable does its own work all right.

Who is not familiar with the electric bell? Who has not used telephone? Who is not enjoying electric light? Electric motors, wireless telegraphy, electric railways, electric modes of constructions and destruction all attest to the great power of electricity; Electricity has developed modern industry and has created many industries. It has enabled man to conquer land, the sea and the air.

Questions:

- (i) Which age has yielded place to the age of electricity? (01)
- (ii) What are the important uses of electricity? (02)
- (iii) Write the meanings of the underlined words. (02)
- (iv) Summarize the passage in your own words and suggest a suitable title. (05)

SECTION – C (Marks 26)

Q. 5 A. Change the narration of any **FOUR** of the following sentences:

(04)

- (i) Akram said to me, "You are not working".
- (ii) "I have washed the clothes", she said.
- (iii) I said to them, "Let us go for a walk".
- (iv) They said to us, "Where had you been living?"
- (v) The teacher said, "Sun sets in the West."

B. Use any FOUR of the following phrasal verbs in sentences:

(04)

(i) Adron with (ii) Look after (iii) Back out
(iv) Get through (v) Pity for

O, 6 Write a letter to your mother who is worried about your health.

(08)

-OR-

Write a letter to the editor of a newspaper commenting on frequent suicidal bomb blasts nowadays.

Q. 7 Write an essay on any ONE of the following topics (150-200 words):

(10)

(i) Courtesy (ii) Life in a Big City

Appendix E**IBCC Conversion Formula**

Ph. No. 9235019
Fax No. 9250454
9250451

GOVERNMENT OF PAKISTAN, MINISTRY OF EDUCATION
INTER BOARD COMMITTEE OF CHAIRMEN

At FBISE Building, H-8/4, Islamabad

No. IBCC/QE/Formula/

31st May, 2010

NOTIFICATION

Inter Board Committee of Chairmen in its 127th meeting, under Resolution No. 23, held on 01-02 March, 2010, at Karachi, approved Equivalence and Revised Conversion Formulae for Foreign Grades, Marks & Scales etc into Pakistani Educational qualification Equivalent Marks.

2. The above resolution shall take effect from 1st May, 2010 for those students who have passed their SSC/HSSC/Diploma/Certificate or equivalent in May 2010 onwards.
3. The resolution of IBCC pertaining to revised formulae is circulated here for information, record and necessary action by all concerned.

Sd/-
(Abdul Qayyum)
Asstt. Secretary

1. All Member of IBCC
2. Chairman, Higher Education Commission, Islamabad
3. President, Pakistan Medical and Dental Council, Islamabad
4. President, Pakistan Engineering Council, Islamabad
5. All Assistant Secretaries of IBCC
6. Programmer IBCC Islamabad to put on it the IBCC Website
7. Assistant Chief, Economic Affairs' Division, Islamabad

GOVERNMENT OF PAKISTAN, MINISTRY OF EDUCATION

(INTER BOARD COMMITTEE OF CHAIRMEN, ISLAMABAD)

Subject: REVISED CONVERSION FORMULAE FOR FOREIGN GRADES/MARKS/SCLAES INTO PAK. EQUIVALENT MARKS AND REVISED FORMULAE WILL BE APPLICABLE ON THOSE APPLICANTS WHO WILL PASS THEIR EXAMINATION (CERTIFICATES/DIPLOMAS) W.E.F. (IN SESSION) 1st MAY, 2010 ONWARD

1. All the Certificates and the grades awarded by the accredited Examination Boards/Authority/Syndicate and institutions recognized by the Education Authority of the country of origin will be considered for equivalence and conversion of marks/grades and will be subjected to the treatment as per approved formula reproduced below:
 - i. The certificates and their marks/grades from accredited/recognized Examination Boards will be equated with Pak equivalent marks as per approved formula.
 - ii. The certificates/transcripts issued by the foreign recognized/accredited institutions/schools/colleges (Internal Examination) will be considered and their grades/marks/scales shall be subjected to 15% deduction to maintain uniformity and range of scores with Pakistani system. As per conversion formula the marks/grades will be first equated with Pak equivalent marks as given in the related context (10%) and then further 5% marks will be deducted from the total making it to 15%.
 - iii. The Revised Conversion Formulae will be enforced and implemented from the session starting from 1st May-2010 and onward.
 - iv. The revised formulae will not be applicable to those who will pass their examination (certificates/Diplomas) before session starting from 1st May, 2010.
 - v. Revised formulae will be applicable on those applicants who will pass their examination (Certificates/diplomas) w.e.f. (in session) 1st May, 2010 onward.
 - vi. The equivalence certificates issued by the IBCC before May-2010 to the candidates or applicants for admission in the professional (Medical/Engineering) Colleges and Universities will not be revised as per approved revised formulae.
 - vii. The actual marks from the recognized examination authority/syndicate/Board will not be considered as such but will be equated to the Pakistani equivalent marks as per approved formulae.

**IBCC CONVERSION REVISED FORMULAE
FOR NUMERIC MARKS:**

No. 1 CONVERSION REVISED FORMULAE FOR NUMERIC MARKS AWARDED BY FOREIGN EXAMINING BODIES.

IN CASE OF EXTERNAL EXAMS:

a) 7.5% deduction be made on numeric marks awarded by Foreign External Examining Bodies all over the world.

IN CASE OF INTERNAL EXAMS :

b) 15% deduction be made on numeric marks awarded by Foreign Accredited/recognized Internal Examining Institutions/Bodies all over the world.

**IBCC CONVERSION REVISED FORMULAE
FOR LETTER GRADES/SCALES**

No.2 GENERAL

Actual scores	Grades	Pak: Equivalent Marks (For External Exams)	Further 5% Deduction	Pak. Equivalent Marks (For Internal Exams)
90 – 100	A	85	5%	80.75
80 - 89	B	75	5%	71.25
70 - 79	C	65	5%	61.75
60 - 69	D	55	5%	52.25
50 - 59	E	45	5%	42.75
40 - 49	F	40	5%	38
35 - 39	G/N/P	35	5%	33.25

No.3 CONVERSION FORMULA FOR IGCSE & GCE LEVELS (BRITISH EXTERNAL EXAMINING BODIES)

IN CASE OF A* IS AWARDED OTHERWISE THE ABOVE GENERAL FORMULA MAY BE IMPOSED.

Grades	Pak. Equivalent Marks (For External Exams)
A*	90 (Maximum Marks)
A	85
B	75
C	65
D	55
E	45
F (before June 2010)	40
G (before June 2010)	35

**NO. 4 FORMULA FOR CAMBRIDGE GRADES (SCHOOL CERTIFICATE)
(FOR OLD SESSIONS)**

Grades	Pak. Equivalent Marks (For External Exams)
A-One	85
A-Two	80

B-Three	75
B-Four	70
C-Five	65
C-Six	60
D-Seven	55
E-Eight	45
N-	35
F	Fail

No. 5 SCOTTISH CERTIFICATE OF EDUCATION

Grades/Scales	Pak. Equivalent Marks (For External Exams)
1	85
2	75
3	65
4	55
5	45
6	40
7	35

Appendix F IBCC Conversion Formula

SES Questionnaire

Dear Students,

Dear Students,

All the information you provide here will be treated in the strictest confidence. Thank you for your valuable time.

1. Personal Information.

2. Academic Information.

a. Are you a student of A Levels or FA/FSc? _____

b. Have you done O Levels or Matric? _____

c. If Matric, how many marks did you obtain in English (class 9 & 10 out of 150)? _____

d. If O Levels, what was your grade in English? _____

3 Information about the family.

3.1 Tick the descriptor which best summarizes the family's situation/status.

3.1. Tick the descriptor which best summarizes the family's vehicle ownership.	
Does your family own a car/van/jeep?	✓
No	
Yes, just one.	
Yes, more than one.	

3.2. Tick the descriptor which best summarizes the family's situation/status.

During the past 12 months, how many times did you travel away on holiday (vacation) with your family?	✓
Not at all	
Once	
Twice	
More than twice	

3.3 Tick the descriptor which best summarizes the family's situation/status.

3.5. Tick the descriptor which best summarizes your family's ownership of computers	
How many computers (including laptops, tablets, etc) does your family own?	✓
None	
One	
Two	
More than two	

3.4. Tick the descriptor which best describes your situation.

3.1. Tick the boxes	Do you have your own bedroom?	✓
Yes		
No		

Section A1: Salient Characteristics of CEFR Levels Chapter 1

Table A1. Salient Characteristics: Interaction & Production (CEFR Section 3.6, simplified)

Table C4: WRITTEN ASSESSMENT CRITERIA GRID

Overall	Range	Coherence	Accuracy	Description	Argument
C2	Shows real flexibility in formulating ideas in different linguistic forms to convey finer shades of meaning precisely to give emphasis and to eliminate ambiguity. Also has a good command of idiomatic expressions and collocations.	Can create coherent and cohesive texts, making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.	Maintains consistent and tight control of grammar, punctuation and spelling, even for non-complex language forms. Errors are rare and non-grammatical errors are few.	Can write clear, smoothly flowing and fully engrossing stories and descriptions of experience in a style appropriate to the genre adopted.	Can produce clear, smoothly flowing, coherently structured reports and essays which present a case or give critical appreciation of proposals, or literary works. Can provide an appropriate and effective logical structure which helps the reader to find significant points.
C1	Has a good command of a broad range of language allowing him/her to select a formulation to express him/herself clearly in an appropriate style on a wide range of general, academic, professional or leisure topics without having to restrict what he/she wants to say. The flexibility in style and tone is somewhat limited.	Can produce clear, smoothly flowing, well-structured text, showing controlled use of organisational patterns, connectors and cohesive devices.	Consistently maintains a high degree of grammatical accuracy, occasional errors in grammar, collocations and idioms.	Can write clear, detailed descriptions and developmental descriptions and imaginative texts in a mostly assessed, personal, natural style appropriate to the reader in mind.	Can write clear, well-structured expositions of complex subjects, underlining the relevant salient issues. Can expand and support points of view at some length with subsidiary points, reasons and relevant examples, and round off with an appropriate conclusion.
B2	Can write clear, detailed official and semi-official texts on a variety of subjects related to his field of interest, synthesising and evaluating information and arguments from a number of sources. Can make a distinction between formal and informal language with occasional less appropriate expressions.	Has a sufficient range of language to be able to give clear descriptions, express viewpoints on more general topics using some complex sentence forms to do so. Language facts, however, expressiveness and idiomativity and use of more complex forms is still stereotypical.	Shows a relatively high degree of grammatical control. Does not make errors which cause misunderstandings.	Can write clear, detailed descriptions of real or imaginary events and experiences in making the relationship between ideas in a clear connected text and following established conventions of the genre concerned.	Can write an essay / report that develops an argument systematically with appropriate highlighting of some significant points and relevant supporting detail. Can evaluate different ideas or solutions to a problem with an essay / report which develops an argument, giving some reasons in support of or against a particular point of view and explaining the advantages and disadvantages of various options.
B1	Can write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence. The texts are understandable but occasional unclear expressions and/or inconsistencies may cause a break-up in reading.	Has enough language to get by with sufficient vocabulary to express him/herself with some circumlocutions on topics such as family, hobbies and interests, work, travel, and current events.	Can link a series of shorter discrete elements into a connected, linear text.	Can write reasonably accurately a repertoire of frequently used "routines" and patterns associated with more common situations. Occasionally makes errors that the reader usually can interpret correctly on the basis of the context.	Can write a short, simple essay on topics of interest.
A2	Can write a series of simple phrases and sentences linked with simple connectors like "and", "but" and "because". Longer texts may contain expressions and show coherence problems which makes the text hard to understand.	Uses basic sentence patterns with reinforced phrases, groups of a few words and formulae in order to communicate limited information mainly in everyday situations	Can link groups of words with simple connectors like "and", "but" and "because".	Uses simple structures correctly, but still systematically makes basic mistakes. Errors may sometimes cause misunderstandings.	Can summarise factual information on a familiar routine and non-routine matters, within his field with some confidence.
A1	Can write simple isolated phrases and sentences. Longer texts contain expressions and show coherence problems which make the text very hard or impossible to understand.	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations.	Can link words or groups of words with very basic linear connectors like "and" and "then".	Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire. Errors may cause misunderstandings.	Can write short simple imaginary biographies and simple poems about people.