

**An Investigation into the Construct Validity of an Academic  
Writing Test in English with Special Reference to the  
Academic Writing Module of the IELTS Test**

Submitted by

Ahmad Abdulrahman Alsagoafi to the University of Exeter  
as a thesis for the degree of Doctor of Philosophy in Education

In February 2013

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

## **Abstract**

The International English Language Testing System (IELTS) is the world's leading high stakes test that assesses the English Language Proficiency of candidates who speak languages other than English and wish to gain entry into universities where English is the language of instruction. Recently, over 3000 institutions in the United States accepted the IELTS test to be an indicator of language proficiency (IELTS, 2012a). Because of this preference for the IELTS test, and its worldwide recognition, there has been an increase in the number of students who are taking the test every year. According to the IELTS website, more than 7000 institutions around the world trust the test results and, not surprisingly, more than 1.7 million candidates take the test every year in one of the 800 recognised test centres across 135 countries (IELTS, 2012a). These candidates include people who seek not only to obtain admission to universities, but also for immigration authorities, employers of certain companies and government agencies.

Acknowledging this popularity and importance to learners of English as a Foreign Language (EFL), this qualitative study has investigated the construct validity of the academic writing module in the IELTS test from the perspectives of the stakeholders (i.e. candidates, lecturers and markers). The aim was to understand why some Saudi students fail to cope with demands of the university despite the fact that they have achieved the minimum requirements in IELTS. In this study, data was collected in two phases in two different settings through open-ended questionnaires, semi-structured observations and semi-structured interviews. Phase I was carried out in the Department of English Language (DEL) at King Faisal University in Saudi Arabia, while Phase II was conducted in one university in the UK. The sample of the study included: 8 students, 6

university lecturers and one marker. In this study, data were analysed and coded into themes by using NVivo 9.

The results of this case study have shown that the stakeholders were doubtful about the issue of readiness of students, which is claimed by IELTS, and they wanted the test to be clearer about how the students were going to cope with university demands upon gaining entry. In addition, with respect to the content validity of the test, this study found that the tasks in the academic writing test to a large extent do not reflect the kind of tasks candidates are likely to encounter at university. Furthermore, this study pointed out that response validity, on the part of students who may not have understood the rubric of the tasks, is another important factor affecting the students' performance. Also, the findings of this study suggested that scoring validity could have a significant effect on the students' scores because of the inconsistency of markers during the scoring process as they may have sometimes failed to assign the students to their corresponding level of proficiency. Consequently, the study provided a set of implications as well as recommendations for future research.

## **Acknowledgements**

All thanks and praise is to ALLAH the Almighty the one who guided me throughout the process of this research or else without His guidance and blessings I could have never completed this dissertation. I am deeply indebted to a number of people for their inspiration, encouragement, cooperation and contribution to the completion of this dissertation. First of all, I would like to express my deepest gratitude and appreciation to my first supervisor Dr. Li Li for her endless encouragement, critical insights, great efforts to explain things clearly and simply, for her full attention and invaluable feedback during the research project. She has never been hesitant to offer her advice to me even when she was my second supervisor. I wish also to express my sincere appreciation to my second supervisor Dr. Susan Jones whose supervision is also enormously valuable and I am deeply touched by her insightful suggestions and helpful feedback, her patient guidance, and her enthusiasm in research. I also owe a debt of thanks and gratitude to Dr. Malcolm MacDonald (Warwick University) who was my former supervisor. When he was my first supervisor, he used to give me support and guidance with respect to the shaping, reading and commenting on the whole manuscript. I owe every step that I have made in this doctoral exploration to all of my supervisors who generously spent numerous hours advising and helping me to come up with a very carefully designed methodology and refine the thesis drafts.

Secondly, my heartfelt thanks go to the academics and friends who helped me throughout my research project. I am really grateful to the university lecturers, students and IELTS examiner who participated in my study. Without their help and support, I would not have

been able to collect information for my study. For the sake of confidentiality, I am not going to mention their names. I would also like to acknowledge the efforts of Assoc. Prof. Keith Postlethwaite (Deputy Head of Graduate School of Education, University of Exeter) and Dr. Dhaif Allah Alzaidi (Assistant Professor, Taiba University) who provided their informative feedback on the face validity of the questionnaires, to Prof. Debra Myhill (Associate Dean, College of Social Sciences and International Studies, University of Exeter) for her insightful comments and sustained encouragement to use the socio-cognitive framework in my study, to Dr. Liu Yongcan (University of Cambridge) with whom I had a fruitful discussion about the Socio-cultural Theory and Activity Theory, to Dr. Anthony Green for his incisive comments during our discussion about the IELTS test, to my mentor Prof. Wendy Robinson (Head of the Graduate School of Education, University of Exeter) for her emotional and academic support during my research project, to Prof. Rüdiger Zimmermann (Philipps University of Marburg) for his reply to my email regarding the models of second language writing, to Prof. Cyril Weir and Lynda Taylor for sharing their knowledge and expertise with me in the Introductory Course to Language Testing in Cambridge university.

I am really very appreciative of a number of academics and experts, especially to those in the field of Language Testing and Education, for their useful comments and conversations in emails or conferences. These include: Prof. Alan Davies (University of Edinburgh), Mr. Alan Addison (IELTS Stakeholder Relations Manager, British Council), Miss Carolyn Walker (Director of INTO Language Centre, University of Exeter), Prof. Charles Alderson (Lancaster University), Miss Chrissi Florides (Director of Studies at

Globe English Centre), Dr. Durrant Philip (Graduate School of Education, University of Exeter), Dr. Guoxing Yu (Graduate School of Education, University of Bristol), Dr. Nasser Mansour (Graduate School of Education, University of Exeter), Dr. Ros Fisher (Director of Doctoral Studies, University of Exeter), and Dr. Vicki Feast (University of South Australia).

Thirdly, my sincere thanks are also due to the Saudi Arabian Cultural Attaché in UK for tracking my progress and encouraging me to complete my PhD, and to my employer King Faisal University (KFU) who granted me this scholarship and offered their full support (academic and financial) for me to collect the data for this study. I am thankful to my supervisor during the data collection process at KFU Dr. Abdullah Alfraidan (Chairman of the Department of English Language) who dedicated his time and effort to make the data collection process successful, and without whom I would not have had the proper access to my participants to collect data. I am also grateful to the staff members in the College of Arts at KFU, especially those in the Department of English Language, who participated in the study.

Finally, my deepest gratitude and love go to my parents who brought me up to make me able to face big challenges. I can never ever forget how sad they were on the day when I left Saudi Arabia to study abroad, and said that “we will rejoice when you pass your PhD with flying colours”. I am truly grateful for all their unconditional love, sacrifices, and prayers. I happily dedicate this PhD to them. My grandmother also had a great positive influence on my progress with her unlimited emotional support and eloquence in saying

the prayers. She used to call me twice a month to make sure that I was progressing well. I thank Allah the Almighty that she is still alive to witness this achievement.

Last but not least, I would like to extend my sincere appreciation and gratitude to my soul mate, my wife Fatimah, who has travelled all the way with me in this long journey. She withdrew from her course in Saudi Arabia as soon as she knew that I was going abroad. I really appreciate this, not to mention her emotional support which helped me to overcome my countless problems that I faced during my research project. I am also thankful for her care of our two lovely children Bandar and Abdullah. Surely, I will be forever thankful for what she did and does for me. I am blessed to be her husband. My special thanks also go to my three brothers, especially Fahad, who were also concerned about making this dream a reality.

## Table of Contents

<b>Acknowledgements.....</b>	<b>4</b>
<b>Abbreviations .....</b>	<b>15</b>
<b>Chapter I.....</b>	<b>20</b>
<b>Introduction.....</b>	<b>20</b>
1.1 Overview of the chapter.....	20
1.2 A Brief background to the IELTS Test.....	20
1.2.1 The development of IELTS.....	23
1.2.1.1 The content and the format of the test.....	24
(a) The listening test .....	25
(b) The reading test .....	25
(c) The writing test .....	26
(d) The speaking test.....	27
1.2.1.2 Marking and assessment.....	27
1.3 Rationale for the study.....	29
1.4 Aim .....	32
1.5 Objectives .....	32
1.6 An overview of the whole dissertation .....	33
<b>Chapter II .....</b>	<b>36</b>
<b>Context of the study .....</b>	<b>36</b>
2.1 Introduction.....	36
2.2 Brief background of the educational system in Saudi Arabia .....	36
a- Pre-School level (2 years).....	37
b- Elementary level (6 years).....	37
c- Intermediate level (3 years) .....	38
d- Secondary level (3 years).....	38
- Regular secondary education.....	38
- Vocational and technical secondary education.....	39
e- Higher education.....	39
2.2.1 The role of assessment in the Saudi Arabian educational system.....	41
2.2.1.1 The General Aptitude Test.....	42
2.2.1.2 The General Aptitude Test for University Graduates.....	42
2.2.1.3 The General Aptitude Test in English.....	42
2.2.1.4 The Achievement Test for Science Colleges (For Men).....	43
2.2.1.5 The Aptitude and Achievement Test for Art Students (For Women).....	43
2.2.1.6 The Aptitude and Achievement Test for Science Students.....	43
2.2.1.7 The Standardised Test of English Proficiency.....	44
2.3 Testing/assessment and/or Educational Organisations Offering Tests in Saudi Arabia .....	44
2.3.1 The British Council in Saudi Arabia.....	44
2.3.2 The Educational Testing Service (ETS) tests in Saudi Arabia .....	45
2.3.3 The Pearson English Tests in Saudi Arabia.....	45
2.3.4 The Graduate Management Admission Council (GMAC) Test .....	45

2.4 The College of Arts .....	46
2.5 A brief description of IELTS candidates in Saudi Arabia.....	46
2.6 Conclusion .....	48
<b>Chapter III.....</b>	<b>50</b>
<b>Literature Review.....</b>	<b>50</b>
3.1 Introduction.....	50
3.2 Theoretical framework.....	51
3.2.1 Writing as a cognitive practice.....	53
3.2.1.1 Models of L1 writing.....	53
(a) Hayes and Flower Model (1980).....	53
(b) Bereiter and Scardamalia Model (1987).....	55
3.2.1.2 Composing processes of second language candidates.....	57
3.2.2 Writing as a social construction practice.....	61
3.2.3 Writing as a social interaction activity.....	64
3.2.4 Writing as a socio-cultural practice .....	66
3.3 A brief introduction to language testing .....	70
3.3.1 Types of tests .....	71
3.3.1.1 Placement tests.....	71
3.3.1.2 Progress tests.....	72
3.3.1.3 Achievement tests.....	72
3.3.1.4 Proficiency tests.....	72
3.3.1.5 Diagnostic tests.....	73
3.4 Essential qualities in language tests.....	74
3.4.1 Validity.....	74
3.4.1.1 Types of validity .....	76
3.4.1.1.1 Content validity.....	79
-Authenticity.....	80
3.4.1.1.2 Response validity.....	85
(a) Macro-planning.....	86
(b) Organisation.....	86
(c) Micro-planning.....	87
(d) Translation.....	87
(e) Monitoring.....	87
(f) Revising.....	88
3.4.1.1.3 Scoring validity.....	89
(a) Criteria/rating scale.....	90
(1) Primary trait scoring.....	90
(2) Holistic scoring.....	92
(3) Analytic scoring.....	93
-Empirical studies on the application of rating scales.....	96
(b) Rater training.....	99
3.5 Empirical studies on IELTS.....	102
3.5.1 Predictive studies on IELTS as a preparation to academic study .....	102
3.5.2 Studies on tasks design for IELTS academic writing .....	104
3.5.3 Studies on assessment of IELTS academic writing tasks .....	109
3.5.4 Impact and washback studies on IELTS and IELTS preparation courses ..	110
3.6 Research Questions.....	116

3.7 Summary and Conclusion.....	117
<b>Chapter IV .....</b>	<b>122</b>
<b>Methodology.....</b>	<b>122</b>
4.1 Introduction.....	122
4.2 Ontology .....	122
4.3 Epistemology .....	124
4.4 Rationale for choosing the qualitative approach .....	126
4.4 Researcher's role .....	128
4.5 Research design .....	134
4.6 Sampling.....	136
4.7 Data collection methods .....	138
4.7.1 Questionnaires.....	139
-Construction of the questionnaires .....	140
4.7.1.1 Students' questionnaires.....	140
a) Phase I.....	141
-Pre-writing test Questionnaire.....	142
-Post-writing test Questionnaire.....	143
b) Phase II.....	143
-IELTS Writing test tasks and University Tasks Questionnaire.....	144
c) Pilot Study.....	144
-Development and validation of the students' questionnaires.....	146
d) Administration of Students' Questionnaires.....	147
4.7.1.2 Marker's Questionnaire.....	148
-The Questionnaire.....	149
-Administration of Marker's Questionnaire.....	149
4.7.2 Interviews.....	150
-Transcription and validation of the interviews.....	151
-Administration of Students' Interviews.....	152
-Administration of Lecturers' Interviews.....	152
4.7.3 Observations.....	153
-Administration of observation.....	153
4.8 Triangulation.....	154
4.9 Type of data .....	155
4.10 Data analysis.....	157
4.10.1 Content analysis.....	157
4.11 Ethical issues .....	158
3.12 Summary of the chapter.....	160
<b>Chapter V.....</b>	<b>162</b>
<b>Findings .....</b>	<b>162</b>
5.1 Introduction.....	162
5.2 Response validity question .....	162
5.2.1 Pre-writing Test Questionnaire .....	163
5.2.1.1 Personal information, prospective course and place of study.....	164
5.2.1.2 General & specific knowledge questions about the AWM in the test...166	166
5.2.2 Post-writing Test Questionnaire.....	173
5.2.2.1 Time allotment.....	174

5.2.2.2	Guidelines and Instructions.....	175
5.2.2.3	Before the test.....	176
5.2.2.4	Difficulty of the questions.....	176
5.2.2.5	Grammar mistakes, style and organisation.....	177
5.2.2.6	Graph/chart/table in task 1.....	178
5.2.2.7	Familiarity with the content of the task.....	179
5.2.2.8	Advantages and disadvantages of the IELTS test.....	181
5.2.3	Observations with students in the mock test.....	182
5.2.4	Interviews with candidates (Phase I).....	183
5.2.4.1	Time to make drafts and complete each task.....	183
5.2.4.2	Task guidelines.....	187
5.2.4.3	Brainstorming.....	190
5.2.4.4	Strategies.....	193
5.2.4.5	Task difficulty.....	198
5.2.4.6	Task order.....	202
5.2.4.7	The questions.....	203
5.2.4.8	Students' insights on the current testing practice of the AWM.....	210
5.3	Content validity question.....	213
5.3.1	IELTS writing test tasks and University Tasks Questionnaire.....	214
5.3.1.1	Students' expectations of university tasks.....	214
5.3.1.2	Similarities & differences between IELTS writing tasks and university tasks.....	215
5.3.1.3	Suitability of IELTS writing tasks for university students.....	217
5.3.2	IELTS Writing test tasks and university tasks Interview.....	218
5.3.2.1	The relationship between IELTS writing tasks and university tasks.....	218
5.3.2.2	The role of IELTS in preparing students to university tasks.....	220
5.3.2.3	Students' perspectives on the level of authenticity of IELTS writing tasks.....	222
5.3.3	Interviews with lecturers (Phase II).....	225
5.3.3.1	Lecturers' perceptions of IELTS tasks.....	226
5.3.3.2	Lecturers' expectations of students who took the IELTS test.....	230
5.4	Scoring validity question:.....	234
5.4.1	Markers' Questionnaire.....	234
5.5	Conclusion.....	237
<b>Chapter VI</b>	.....	<b>239</b>
<b>Discussion</b>	.....	<b>239</b>
6.1	Introduction.....	239
6.2	The key findings of the study.....	239
6.2.1	Phase I.....	240
6.2.1.1	Key findings for response validity question.....	240
6.2.1.2	Key findings for scoring validity question.....	241
6.2.2	Phase II.....	242
6.2.2.1	Key findings for content validity question.....	242
6.3	The stakeholders' perceptions about the test.....	243
6.3.1	The students' perceptions of the value of the test.....	243
6.3.2	The stakeholders' perceptions of their test awareness.....	244

6.3.3 Stakeholders' doubts about IELTS correspondence to university tasks .....	250
6.3.4 Stakeholders' doubts about IELTS as a measure .....	255
6.4 The students' engagement with the IELTS writing test .....	263
6.4.1 The students' perceptions of test difficulty and duration .....	263
6.4.2 The students' perceptions of test prompts and instructions .....	266
6.4.3 The perceived test taking (Students) and scoring strategies (Markers) .....	268
6.4.4 Expertise in second language (L2) writing .....	273
6.5 The connection between perceptions and engagement .....	274
6.5.1 Change of perspectives .....	274
6.6 Summary of the chapter .....	277
<b>Chapter VII .....</b>	<b>281</b>
<b>Conclusion:.....</b>	<b>281</b>
7.1 Introduction.....	281
7.2 Summary of the study.....	281
7.2.1 The stakeholders' perceptions about the mismatch between IELTS tasks and university tasks.....	282
7.2.2 Stakeholders' doubts about IELTS as a measure.....	284
7.2.3 Tracking the change of perspectives .....	287
7.3 Significance of the study.....	288
7.4 Limitations of the study .....	290
7.5 Implications of the study .....	291
7.5.1 Stakeholders.....	291
7.5.2 Test providers.....	293
7.5.3 Methodological implications of the study.....	295
7.6 Recommendations for future research .....	296
7.7 Final conclusion to the study .....	298
<b>References .....</b>	<b>300</b>
<b>Appendices .....</b>	<b>335</b>
<b>Appendix I.....</b>	<b>336</b>
<b>Data Collection Tools &amp; Key Findings Of The Study.....</b>	<b>336</b>
Phase I.....	336
IELTS Awareness Form for Candidates (Pre-Writing Test Phase I).....	336
IELTS Awareness Form for Candidates (Post-Writing Test Phase I) .....	338
IELTS Post-Writing Test Interview (Candidates) .....	340
IELTS Questionnaire for Markers .....	342
Phase II .....	344
IELTS Tasks and University Tasks Form for Candidates (Phase II).....	344
Interview Questions for Candidates.....	345
Interview Questions for Lecturers.....	346
Sample of a transcription of one interview .....	347
Key Findings (according to data collection tools).....	352
Timeframe of Research.....	361
Consent Form.....	362
<b>Appendix II .....</b>	<b>363</b>
<b>IELTS Tasks &amp; Scoring Guides.....</b>	<b>363</b>
Sample of IELTS Writing Task 1 .....	363

Model answer for IELTS Writing Task 1 .....	364
Sample of IELTS Writing Task 2 .....	365
Model answer for IELTS Writing Task 2 .....	366
Scoring Criteria for Task One .....	367
Scoring Criteria for Task Two .....	369
<b>Appendix III.....</b>	<b>371</b>
<b>Weir’s Socio-cognitive Framework for Writing.....</b>	<b>371</b>
<b>Appendix IV .....</b>	<b>372</b>
<b>NVivo Samples.....</b>	<b>372</b>
<b>Appendix V.....</b>	<b>374</b>
<b>Mock IELTS Test Observation Schedule.....</b>	<b>374</b>

### List of Tables

Table 1: History of the IELTS Test .....	22
Table 2: IELTS Band Scores .....	28
Table 3: The educational ladder in Saudi Arabia .....	40
Table 4: Summary of empirical research on writing processes of L2 learners.....	61
Table 5: Task Characteristics.....	82
Table 6: A comparison of holistic & analytic scales on 6 qualities of test usefulness .....	94
Table 7: Tasks and Skills in IELTS versus those in Research Degrees.....	106
Table 8: Summary of research paradigm (ontology, epistemology & methodology), design & methods .....	134
Table 9: Participants, context of the study, time and duration.....	138
Table 10: Framework of the study .....	154
Table 11: An overview of questions, participants and methods.....	156
Table 12: Personal information, prospective course and place of study.....	165
Table 13: Students’ knowledge of the test.....	169
Table 14: Advantages and Disadvantages of the IELTS test.....	181
Table 15: Summary of the students’ views on the similarities and differences between IELTS tasks and University tasks.....	216

**List of Figures**

Figure 1: Cognitive validity parameters in writing .....86  
Figure 2: Interactions in performance assessment of speaking skills.....99

**List of Diagrams**

Diagram 1: Test Structure ..... 24  
Diagram 2: Theoretical Framework of the Study ..... 52

## **Abbreviations**

<b>ACC</b>	Accounting
<b>AMIDEAST</b>	America-Mideast Educational and Training Services
<b>AppL</b>	Applied Linguistics
<b>AR</b>	Arabic
<b>Aus</b>	Australia
<b>AWM</b>	Academic Writing Module
<b>BEC</b>	Business English Certificate
<b>BERA</b>	British Educational Research Association
<b>BULATS</b>	Business Language Testing Service
<b>CB</b>	Computer Based
<b>DEL</b>	Department of English Language
<b>EAP</b>	English for Academic Purposes
<b>Edu</b>	Education
<b>EFL</b>	English as a Foreign Language
<b>ELTS</b>	English Language Testing Service

<b>EN</b>	English
<b>EPTB</b>	English Proficiency Test Battery
<b>ESL</b>	English as a Second Language
<b>ESOL</b>	English for Speakers of Other Languages
<b>ETS</b>	Educational Testing Service
<b>F</b>	Female
<b>FN</b>	Finance
<b>G</b>	Graduate
<b>GMAC</b>	Graduate Management Admission Council
<b>GMAT</b>	Graduate Management Admission Test
<b>GPA</b>	Grade Point Average
<b>GRE</b>	Graduate Record Examination
<b>iBT</b>	TOEFL Internet Based Test
<b>IDP</b>	International Development Programme
<b>IELTS</b>	International English Language Testing System
<b>ITP</b>	Institutional Testing Program
<b>KFU</b>	King Faisal University

<b>L1</b>	First Language
<b>L2</b>	Second language
<b>M</b>	Male
<b>NCAHE</b>	The National Centre for Assessment in Higher Education
<b>PBT</b>	Paper Based Test
<b>PG</b>	Postgraduate
<b>PTE</b>	Pearson Test of English
<b>SCT</b>	Sociocultural Theory
<b>SE</b>	Special Education
<b>TAFE</b>	Technical And Further Education
<b>TESOL</b>	Teaching English to Speakers of Other Languages
<b>TLU</b>	Target Language Use
<b>TOEFL</b>	Test Of English as a Foreign Language
<b>TOEIC</b>	Test Of English for International Communication
<b>UCLES</b>	Cambridge Local Examination Syndicate
<b>UG</b>	Undergraduate
<b>UK</b>	United Kingdom

**US** United States

**W** Withdrawn

# CHAPTER 1

## **Chapter I**

### **Introduction**

#### **1.1 Overview of the chapter**

The aim of this chapter is to introduce this study. It starts by giving a brief background about the IELTS test, namely, its content, format, marking system and how it was developed. This chapter also explains the rationale for doing the study, the aim of the study and objectives. It concludes with a section about the organisation of the chapters in this dissertation.

#### **1.2 A Brief background to the IELTS Test**

The International English Language Testing System (IELTS) is a high-stakes test which was launched in 1989 to assess the English language proficiency of international students who want to join universities in English speaking countries. This test is the result of numerous revisions of the earlier version (See Table 1), the English Language Testing Service (ELTS), which was introduced in 1980 (Alderson, 1991). The ELTS test replaced the English Proficiency Test Battery (EPTB), which was introduced in 1965 by the British Council to measure the language proficiency of international applicants to UK universities and colleges (IELTS, 2012b). Currently, the English for Speakers of Other Languages (ESOL) Examinations at the University of Cambridge, the British Council and the International Development Programme (IDP): IELTS Australia jointly own the IELTS test (IELTS, 2012c).

Now, countries such as the United Kingdom, which has increasingly become an attraction to many overseas students, together with Australia, New Zealand and Canada, prefer applicants who hold IELTS test reports, despite the fact that there are other tests available, such as the Test of English as a Foreign Language (TOEFL). This can be clearly seen in the application forms for admission into universities, with the IELTS being the first choice for university admission and for admission to the university preparation courses in language centres in these countries.

Recently, over 3000 institutions in the United States accepted the IELTS test to be an indicator of language proficiency (IELTS, 2012a). Because of this preference for the IELTS test, and its worldwide recognition, there has been an increase in the number of students who are taking the test every year. According to the IELTS website, more than 7000 institutions around the world trust the test results and, more than 1.7 million candidates take the test every year in one of the 800 recognised test centres across 135 countries (IELTS, 2012a). These candidates include people who seek not only to obtain admission to universities, but also for immigration authorities, employers of certain companies and government agencies.

Moreover, this growing interest in the IELTS test resulted in the launching of more preparation programmes and the publishing of more specimen materials around the world. For instance, the British Council offers preparation courses for the IELTS test and a specimen materials booklet (published by Cambridge University) for those who do not wish to enrol in the preparation course. There are also numerous preparation courses and materials provided by private institutions and language centres in most universities. The purpose of such preparatory courses and self-study kits, which can sometimes be online

materials, is to familiarise and prepare the candidates for the actual test. The following section sheds light on the developments of the IELTS test and its content.

Test	Launch Date/ Revisions	Content	Type of Questions & Approach	Duration
EPTB	1965	<b>A. Listening</b> 1. Phonemic discrimination (in isolation) 2. Phonemic discrimination (in context) 3. Intonation and stress (in conversation) 4. Comprehension (science and non-science texts) <b>B. Reading</b> 5. Grammatical structure 6. Reading speed 7. Comprehension (science and non-science texts)	-Multiple Choice Questions  -Structuralism	1hr 45mins
ELTS	1980	<b>Modules:</b> 1. Life Sciences 2. Social Sciences 3. Physical Sciences 4. Technology 5. Medicine 6. General Academic <b>Tests:</b> 2 General tests: Reading & Listening 3 Specific tests: Study Skills, Writing & Interview.	Multiple Choice Questions & Writing & Speaking tests with thematically linked subjects  -Communicative Competence	Approx.3hrs
IELTS	1989	<b>Modules:</b> A- Physical Science and Technology B- Life and Medical Sciences C- Business Studies and Social Sciences <b>Tests:</b> Reading Module A Module B Module C General Writing Module A Module B Module C General Listening Non-specialised Module Speaking Non-specialised Module	Multiple Choice Questions & Writing & Speaking tests with integrated skills  -Communicative Competence	Approx. 2hrs 25mins.
IELTS	Revision 1995 till now	<b>Modules:</b> A- General B- Academic <b>Tests:</b> All candidates take the same listening and speaking tests with difference in the academic/general reading and writing tests.	Multiple Choice Questions & Writing & Speaking tests with no thematic link between the skills  -Communicative Competence	Approx. 2hrs 45 mins.

Table 1: History of the IELTS Test

### **1.2.1 The development of IELTS**

The IELTS test developed over time through revision projects commissioned by Cambridge Assessment, through internal and external research. It has emerged as a result of the ELTS test revision plan that took place from 1987 until 1989. It was recommended in this revision plan that the number of subject-specific modules should be reduced from six to four, the Non-Academic Module needed to be replaced by the General Training Module and the word “International” had to be attached to it to acknowledge the involvement of the IDP who had just joined the British Council and the UCLES in administering the test (Charge and Taylor, 1997).

After 1989, when the IELTS test went live, its specifications were first revised in 1993, in a process that was “designed specifically to validate the test, evaluate the impact of the test, provide relevant information to test users and to ensure that a high quality of service is maintained” (IELTS Revision Specifications, 1993, p.64). However, the major change in IELTS was, with no doubt, made in the 1995 revision. The purpose of this revision, based on recent developments and research on applied linguistics and language testing, was to address four factors: practical concerns, administrative problems, technological developments and theoretical issues (Charge and Taylor, 1997). Since then, minor revisions were undertaken at the level of skills, e.g. speaking revision in 2001 and writing revision in 2005, and a new format of the test like the Computer Based (CB) IELTS test was introduced in 2005 (Davies, 2008).

### 1.2.1.1 The content and the format of the test

The test has two modules, namely General (for courses and programmes at below degree level, for migration and employment purposes) and Academic (for postgraduate or undergraduate courses and enrolment at a professional level), depending on the nature of the course in which candidates will be enrolled. The test is comprehensive, in terms of measurement and coverage, to the extent that it allows for the four skills, i.e. listening, speaking, reading and writing, to be tested separately.

All candidates take the same speaking and listening test with the difference only being manifested in the reading and writing tests. The listening, reading and writing tests have to be taken on the same day (either paper and pencil or on-screen), while the speaking test can be taken directly after these tests, if possible, or within 7 days. The duration of the test is 2 hours and 45 minutes. Diagram 1 below shows how the total test is distributed across the four skills. Each of these tests will be discussed in detail in the following sections and ordered in the same pattern that they appear in the test.



Diagram 1: Test Structure. (Obtained from IELTS Website, 2009)

### **(a) The listening test**

Candidates in this test listen to a conversation, monologue or lecture which is played once only. After that, candidates are given ten minutes to transfer their answers from the question paper into the answer sheet. Most frequently used question types are: multiple choice, short answer, sentence completion, notes, table, summary flow chart or diagram completion, classification and labelling of a diagram that has numbered parts. Instructions on how to answer these questions are given at the beginning of each section.. The level of difficulty increases as candidates progress through the test.

### **(b) The reading test**

In this test, candidates are given a question booklet and an answer sheet where they should transfer their answers. The types of questions after each reading passage are multiple choice, sentence completion, short answer, yes/no and true/false, matching, questions that require scanning to identify specific information, choosing headings for paragraphs or sections of a text, notes, table, summary flow chart or diagram completion, classification and labelling of a diagram that has numbered parts. The topics of reading passages are of general interest to candidates with an increasing level of difficulty. The general module includes three texts (one being descriptive) which candidates are likely to encounter on a daily basis in an English speaking country, whereas the academic module is composed of three texts (one being argumentative) which are somehow similar to the kind of texts candidates are expected to deal with at university level.

### **(c) The writing test**

This is the third test candidates take on the test day which lasts for 60 minutes. In the academic module, the test consists of two tasks in which students are asked to write in formal style of English, in an answer booklet, a descriptive report and produce an argumentative essay. More specifically, the first task requires students to describe and interpret data in a chart, table, graph, diagram or flow chart. They need to write 150 words in about 20 minutes. Assessment is based on the student's ability to organise and compare data, describe something or event, and explain how something works (IELTS Handbook, 2007). In contrast, in the second task students need to argue an issue, determine their position either for or against a proposition, present and then justify their opinions by providing evidence or supporting information for each point related to the topic of discussion. The students need to write 250 words in about 40 minutes. They are assessed on their ability to solve a problem, present and justify an opinion, compare and contrast evidence and evaluate and challenge ideas (IELTS Handbook, 2007). The topics are general and suitable for candidates wishing to study at undergraduate or postgraduate courses at university. In the general writing module, candidates, firstly, write a letter which is slightly more personal than task 2 to ask for information or to explain a situation in a 150 words letter for about 20 minutes; secondly, they write a short essay in which they show their ability to present a position, construct an argument and discuss ideas in a 250 words essay for about 40 minutes (IELTS, 2009). In the first task students are assessed on their ability to engage in personal correspondence, provide general factual information and express opinions whereas in the second task assessment is based on their ability to outline a problem and provide a solution, justify an opinion, provide general

factual information and evaluate evidence or an argument (IELTS Handbook, 2007). The topics are general and suitable for any programme of study.

#### **(d) The speaking test**

This is the last test candidates take in IELTS. All candidates regardless of their module type (general or academic), take the same speaking test. The test lasts for 15 minutes and is in the form of an interview. This interview requires the candidate to talk about some aspects of his/her past, present and future situations. The examiner also asks the candidate about issues of general interest. The whole interview is recorded for administrative purposes. The following section discusses the marking and assessment of the test.

##### **1.2.1.2 Marking and assessment**

Candidates in the IELTS test are assessed on a 9-band scale (see Table 2) where they “receive an overall score from 1-9 together with a score for each skill module” (Davies, 2008, p.105). In the case of writing, which is the focus of this study, examiners assess written performance by using a set of performance descriptors at each of the 9 bands.

More specifically, they “award a band score for each of four criterion areas: Task Achievement (for Task 1), Task Response (for Task 2), Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy” (IELTS, 2010a) (See Appendix II).

Students will be penalised if they fail to achieve the required word count, especially with task 2, which carries twice as much to the writing score, if their response is off-topic, if they copy directly from the question paper or if their writing is plagiarised (Cambridge

ESOL, 2011). Likewise, those who write more than the required limit may produce irrelevant responses and they may not have enough time to complete the other task.

<b>9</b>	<b>Expert user</b>	Has fully operational command of the language: appropriate, accurate and fluent with complete understanding.
<b>8</b>	<b>Very good user</b>	Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriate words. Misunderstandings may occur in unfamiliar situations. Handles complex detailed argumentation well.
<b>7</b>	<b>Good user</b>	Has operational command of the language, though with occasional inaccuracies, inappropriate words and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.
<b>6</b>	<b>Competent user</b>	Has generally effective command of the language despite some inaccuracies, inappropriate words and misunderstandings. Can use and understand fairly complex language particularly in familiar situations.
<b>5</b>	<b>Modest user</b>	Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.
<b>4</b>	<b>Limited user</b>	Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.
<b>3</b>	<b>Extremely limited user</b>	Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.
<b>2</b>	<b>Intermittent user</b>	No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty understanding spoken and written English.
<b>1</b>	<b>Non user</b>	Essentially has no ability to use the language beyond possibly a few isolated words.
<b>0</b>	<b>Did not attempt the test</b>	No assessable information provided.

Table 2: IELTS Band Scores (Obtained from IELTS website, 2012e)

### **1.3 Rationale for the study**

As an overseas student whose native language is not English, I had to take the test in 2004 to gain entry into an English speaking university. At that time, IELTS was just introduced in Saudi Arabia through the British Council and I did not have the slightest idea of what the test was going to be like until I visited the IELTS website and bought the Specimen Materials Booklet to prepare for the test. I found that the test was comprehensive in terms of measuring the four skills which some proficiency tests like the TOEFL test ignored (the speaking skill was not assessed at that time).

Moreover, my interest grew further as the test started to attract global attention and increase the students' intake in spite of on-going competition with other tests. It is now gaining ground over the TOEFL test as it is accepted in more than 3000 institutions in the United States and it could be true that 'the world speaks IELTS' (IELTS, 2012a). Indeed, the fact that it had achieved such a high status among other functioning language tests, and the dynamic process of its continuing development to render it a more efficient assessment tool of the English language proficiency of non-native speakers, is what made the IELTS test the focus of my interest.

More specifically, the IELTS test consists of two modules, namely, general and academic. I decided to study the academic module and chose writing from that module. The reason for choosing academic writing and not any other skill in the IELTS test arose from the fact that assessment of students' performance at university level is based mainly on what they write. They are expected to write research papers or dissertations where the language is characterised as being highly academic. It is not surprising that universities acknowledge the importance of this skill through the demand for a specific band score on

the writing test on IELTS. This band score ranges from 5 to sometimes 7 according to the type of course in which the candidate will be enrolled. Universities very often review the minimum of this writing band score and may raise it from time to time to make sure that candidates are able to meet the expectations of the target course of instruction. Such a decision may be based on the fact that universities have discovered that candidates' writing is inadequate for a tertiary programme of study. However, it seems that raising the IELTS band score could be practical, but may not be a preferable (Wette, 2012), solution in the long run to those students who fail to cope with the demands of the university despite the fact that they have achieved the minimum entry requirements through the IELTS test.

On the other hand, since 2005 the Saudi government has started sending its students abroad to continue their education in many countries around the world. Most of these students were sent to English speaking countries. English in Saudi Arabia is taught as a Foreign Language (EFL). Therefore, students had little exposure to the language in their country. They had to show evidence through tests such as TOEFL or IELTS that they have sufficient English language proficiency to undertake university tasks in the target contexts. Candidates who apply to UK or Australian universities mostly sit for the IELTS test. Since 2005 when my scholarship started, I met many students in Australia where I did my Masters Degree, and UK where I am at the moment. Those students were complaining that the IELTS test did not prepare them for university and there was little correspondence between the IELTS tasks and university tasks. They added that being unprepared for university could be one reason for their failure. Despite this limitation, the Saudi students who were seeking admission in UK or Australian universities were under

pressure to take the IELTS test in order to get an unconditional offer for the course in which they wanted to enrol. Later on, getting the required score in IELTS or any similar test was a condition by the Saudi government for those students who wanted to get a scholarship. Based on the above concerns, I felt that there is a compelling need to have a full understanding of the cause of failure of some Saudi students at university despite the fact that they have achieved the minimum requirements in IELTS.

After reviewing the pertinent literature regarding the validity of the writing module in the IELTS test, I noticed that most studies have been fragmented in terms of addressing one type of validity (i.e. content, response or scoring), or sometimes one writing task with a focus on process or product (e.g. Banerjee, 2000; Mickan et al., 2000; Rignall and Furneaux, 2002; Mickan, 2003; Mickan and Slater, 2003; Mayor et al., 2003; Mayor et al., 2007; Moore and Morton, 2007; O'Loughlin and Wigglesworth, 2007; and Yu et al., 2011). Another motivation behind the current study is that "it is important to provide context specific studies which may contribute to the wider literature of IELTS validity" (Woodrow, 2006, p.52).

More specifically, "more research is necessary" in the IELTS writing test "especially in the areas of raters, scale, task, test taker behaviour, and topic comparability to diagnose and minimize sources of error in testing writing" (Uysal, 2010, p.3). In addition, "future research should be performed to explore whether the characteristics of the IELTS test tasks and the TLU tasks match" (ibid, p.6). Recently, Yu et al. (2011) noted that there was a dearth of research with respect to the test taker's cognitive performance when completing task one in IELTS writing test. Hence, to the best knowledge of the researcher, there is no study about the academic writing module of the IELTS test that

looked at both of the writing tasks from a socio-cognitive perspective and investigated their construct validity (i.e. content, response and scoring) from the perspectives of the stakeholders in a case study that followed the same participants on two occasions. More specifically, this study argues that the test does not reflect the kind of tasks candidates are likely to encounter at university. Arguably, that is why many candidates fail to cope with the demands of university study. In addition, it contends that it may not only be the problem of the construction of the test but may also be due to students who may not have responded appropriately to the test tasks, or to examiners who may have failed to allocate students to the appropriate level of proficiency due to their inconsistency during the scoring process.

#### **1.4 Aim**

- To investigate the construct validity of the academic writing module in the IELTS test from the standpoint of the stakeholders in order to understand the reason/s for the failure of Saudi students at university.

#### **1.5 Objectives**

- 1- To examine the relationship between the IELTS writing test tasks and university tasks in terms of length, style, level of difficulty, balance of themes, suitability to the learners and authenticity.
- 2- To describe the process of making sense of a task and the use of different genres in different contexts.

3- To investigate the consistency of using the rating scales by IELTS examiners.

### **1.6 An overview of the whole dissertation**

This dissertation is organised into seven chapters. **Chapter I** gives a brief introduction to the IELTS test (i.e. its content, format and development). This chapter also sheds some light on the rationale for doing the study, the aim of the study and objectives. **Chapter II** provides a brief description of the context of the study with a general overview of the educational system in Saudi Arabia.

**Chapter III** reviews previous literature with respect to writing composition (process & product) and assessment. It looks at writing composition from a socio-cognitive perspective. It gives a brief introduction to the field of language testing, and validity. It then lists the types of tests and types of validity. After discussing the theoretical construct, the review shifts to the practical side with special reference to empirical studies on IELTS related to the types of validity being investigated in this study. Finally, the research questions for this study are presented. **Chapter IV** describes the methodology adopted in this study, the philosophical assumptions underlying the current study, and the rationale for choosing the research approach. It then explains the researcher's role, the method of sampling, the context and the participants of the study. The chapter sheds some light on data analysis, and ethical issues.

**Chapter V** reports the findings of Phase I and Phase II of the study. These findings are all qualitative and were gained from questionnaires, interviews, and observation. The results of this study were analysed by using NVivo (v 9.2.81.0) and will be presented according to the research questions for this study. **Chapter VI** aims at exploring the links

between the phenomena investigated in the study and the main themes that arose during the analysis. The findings will be reviewed and summarised in order to provide answers for the research questions and to show whether there is a consistency or inconsistency between the results of the study and similar studies in the literature. The discussion of the findings will be presented thematically.

**Chapter VII**, which is the conclusion chapter of this study, will focus on transforming the stakeholders' perceptions into statements that describe their importance in informing the current assessment practice of the academic writing module in the IELTS test. In other words, it attempts to demonstrate the contribution of the study to the literature around language testing in general and performance assessment of academic proficiency in second language writing through IELTS in particular. This chapter will also draw our attention to the limitations of the study and will propose some implications and recommendations for future research.

# CHAPTER 2

## **Chapter II**

### **Context of the study**

#### **2.1 Introduction**

The aim of this chapter is to give a brief background about the context of the study. It starts with a general overview about the educational system in Saudi Arabia and lists the levels of education that the Saudi government provides for the students. It then sheds some light on the local assessment practices common in Saudi Arabia. This is followed by a section about the foreign organisations that offer tests in Saudi Arabia. This chapter ends with a short description of the specific place of the study and the IELTS candidates in Saudi Arabia.

#### **2.2 Brief background of the educational system in Saudi Arabia**

According to Article 50 of the Educational Policy in the Kingdom of Saudi Arabia, Saudi students should be taught at least one foreign language along with their native language (Ministry of Education, 2012). The document explained that this foreign language must be one of the modern languages, i.e. a language in use today so that Saudi students can be able to communicate with people from different countries and cultures, in order to gain knowledge and get acquainted with the latest developments in all relevant fields of knowledge. Actually, this article expounds the rationale for teaching English in Saudi schools. It also highlights the importance of English in the daily lives of Saudi citizens as being the only foreign language taught in Saudi schools and the medium of instruction on some courses at Saudi universities. More specifically, this section will describe briefly

the educational system in Saudi Arabia and highlight the strong presence of English in most of its levels (see Table 3).

**a- Pre-School level (2 years)**

At this level, children are prepared for elementary education. It lasts for 2 years in total. Boys and girls are enrolled at this level at the age of four in the nursery and at the age of five in the preliminary school.

**b- Elementary level (6 years)**

This level is compulsory for all students because it is considered as the basis of the education system. Children can start studying at this level at the age of six (children who are three months younger, i.e. 5 years and nine months can be exempted from this stipulation). Students spend six years at this level with two semesters in each year. Grade 1-4 students are assessed formatively throughout the year, while Grade 5-6 students are assessed summatively (mid-term and final term exams) in a two-week period. The duration of the class is 45 minutes. Boys and girls are segregated from each other but there is a standard curriculum for both of them. Since 2005, English has been taught in the sixth grade. Actually, it is now under trial to be taught in the fourth Grade in some schools. If this proves to be successful, it will be officially introduced to all public schools in the kingdom. Currently, private education offers English classes as early as the kindergarten and at every stage in their schools.

### **c- Intermediate level (3 years)**

As soon as students complete the elementary level successfully, they are encouraged to continue their education at the intermediate level. At this level, students have become almost 12 years old. They are assessed summatively through mid-term and final-term exams in each semester. English is taught in all of the three stages. The students have to study English for 4 days a week in a 45-minute class.

### **d- Secondary level (3 years)**

Upon completing the intermediate level successfully, the students at the age of 15 to 19 year-old spend three years at the secondary level in either regular or vocational and technical secondary education. They are described in some detail below as follows:

#### **- Regular secondary education**

This level consists of three years. The first year offers a general curriculum for all of the students. With respect to the remaining two years, the students can choose one of these three routes: Administration & Social Science, Natural Science, Shariah & Arabic Studies. There are two semesters in each year and the class is 45-minute long.

Assessment is based on a mid-term and a final exam at the end of the semester. In order to get the Secondary School Certificate, the students must pass each subject with at least 50% of the maximum score. English, at this level, is still a compulsory subject for all students enrolling in different majors.

### **- Vocational and technical secondary education**

Graduates from intermediate schools have the opportunity to choose another route for learning, namely the vocational and technical education. The duration of this secondary level is three years. In vocational education students are trained to handle highly sophisticated jobs at major companies. Likewise, the technical education prepares students to be highly skilled in certain areas, which the country needs to keep abreast with the latest changes in technological developments taking place around the world. Technical education is divided into three majors: Industrial, Commercial, and Agricultural.

### **e- Higher education**

The post secondary system of education in Saudi Arabia is a separate higher education programme where students spend 4 to 5 years in order to get a Bachelor degree in a particular subject. The majority of universities in Saudi Arabia are public with few but rapidly growing private ones. Currently, there are 24 public universities and 10 private universities in the Kingdom (Ministry of Higher Education, 2012) offering Diploma and Bachelor degrees. Most of these universities provide Master and PhD degrees.

English is taught at the bachelor degree for all students. It is provided as a general subject in order to give students basic knowledge of the language needed for survival at university. The students study one or more English subjects to improve their basic reading and writing skills that they have learned in the introductory course earlier.

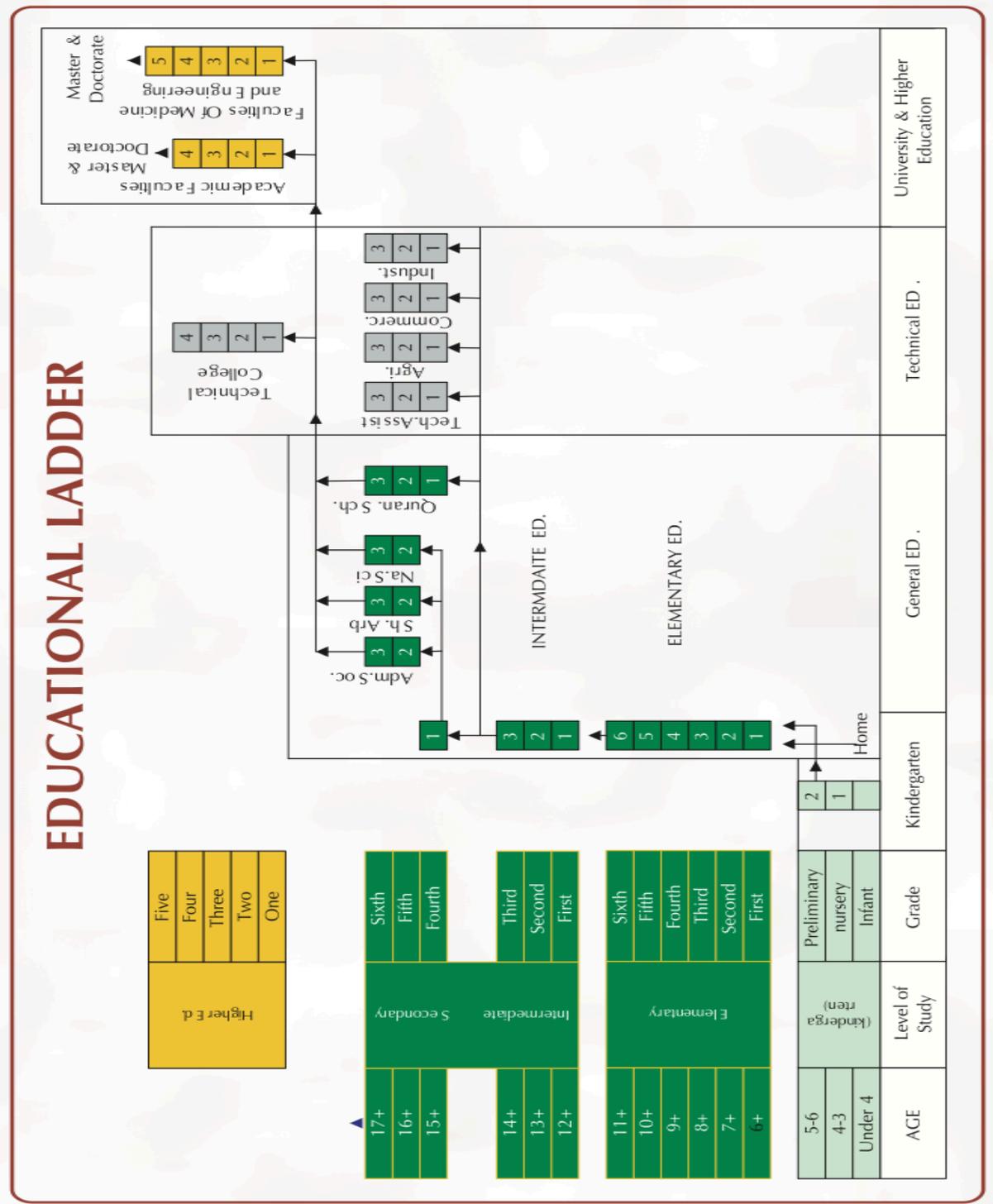


Table 3: The educational ladder in Saudi Arabia (retrieved from the Saudi Arabian Cultural Mission to the US, 2012)

### **2.2.1 The role of assessment in the Saudi Arabian educational system**

Assessment plays an important role in the educational system in Saudi Arabia. Students are assessed at almost all ages through achievement tests. These tests can be in the form of formative assessment where an immediate feedback about the students' performance is sought. This is done frequently in primary schools through quizzes, question and answer sessions, homework, observations, activities, etc. On the other hand, the rest of the students at other levels until higher education level are mainly assessed through summative tests at the end of the year or a period of instruction, with a focus sometimes on formative tests to check the students' progress. These summative tests can be in the form of mid-term exams, final exams, or projects and term papers for university students. Evidently, assessment is an important component of the education system in Saudi Arabia.

There are several other types of tests that the students need to take after finishing the secondary school according to the programme they wish to join. The purpose of this review of tests here is to draw the attention of the reader to the current growing testing practices taking place in Saudi Arabia and to acknowledge the fact that "...testing has become big business" (Spolsky, 2008, p.297).

The National Centre for Assessment in Higher Education (NCAHE) offers a number of tests for Saudi students. These include:

- 1- The General Aptitude Test
- 2- The General Aptitude Test for University Graduates
- 3- The General Aptitude Test in English

- 4- The Achievement Test for Science Colleges (For Men)
- 5- The Aptitude and Achievement Test for Art Students (For Women)
- 6- The Aptitude and Achievement Test for Science Students (For Women)
- 7- The Proficiency Test in English

The following is a detailed description of these tests (The National Centre for Assessment in Higher Education, 2012):

#### **2.2.1.1 The General Aptitude Test**

This test is considered to be an admission test to universities and higher education institutions. Therefore, all students who wish to enroll in universities and higher education institutions have to take this test. It is a standardised test written in Arabic language. The duration of the test is two and a half hours. The test measures the analytical as well as the reasoning abilities of students. It is divided into two parts: Language and Mathematics.

#### **2.2.1.2 The General Aptitude Test for University Graduates**

This test is specially designed for university graduates who are willing to continue their education to do a Master Degree or any higher education qualification. This test is similar to the Graduate Record Examination test (GRE). It consists of three parts: Language, Mathematics and Logical and Analytical Reasoning.

#### **2.2.1.3 The General Aptitude Test in English**

This test is similar to the General Aptitude Test which is written in Arabic, but it is not exactly the same, i.e. it is not translated literally into English. As the General Aptitude

Test, this test measures the analytical as well as the reasoning abilities of students in two sections: Language and Mathematics. The duration of the test is roughly three hours including the instructions given to students at the beginning of the test.

#### **2.2.1.4 The Achievement Test for Science Colleges (For Men)**

This test covers most of the general concepts in these subjects: Biology, Chemistry, Physics, Mathematics, and English. The content of the test is derived from these subjects that the student took during the three years they spent in the secondary school. The test is divided into five sections according to the above subjects with a 20% given to each of them. The test lasts for two and a half hours.

#### **2.2.1.5 The Aptitude and Achievement Test for Art Students (For Women)**

This test is for women students who are specialising in Arts and would like to join Arts or Humanities colleges at certain universities. The test is divided into six sections: two sections for the General Aptitude Test and four sections for the Achievement Test. The duration of the test is three hours as each section is given 30 minutes.

#### **2.2.1.6 The Aptitude and Achievement Test for Science Students (For Women)**

This test is for women students who are specialising in Science and would like to join any college at certain universities. The test is divided into six sections: two sections for the General Aptitude Test and four sections for the Achievement Test. The duration of the test is three hours as each section is given 30 minutes.

### **2.2.1.7 The Standardised Test of English Proficiency**

The purpose of this test is to measure the proficiency of students in English language. The test consists of 100 multiple-choice questions (40% Reading Comprehension, 30% structure, 20% Listening Comprehension and 10% Writing Analysis). Successful candidates are informed about their level of proficiency in order to seek admission into English departments, or to be exempted from studying English subjects at university or somewhere else where English is needed for communication.

To sum up, one might be able to infer that language assessment in the Saudi Arabian context is growing rapidly. In addition to the above tests introduced recently by the NCAHE, almost all universities in Saudi Arabia have entrance exams, placement tests or sometimes interviews for the programmes that they offer. Once the students pass these tests or interviews they are placed/accepted into the programme of their study.

## **2.3 Testing/assessment and/or Educational Organisations Offering Tests in Saudi Arabia**

This review below highlights the organisations that offer major tests in Saudi Arabia.

### **2.3.1 The British Council in Saudi Arabia**

The British Council offers a wide range of tests in order to measure the general and academic language proficiency in English. For example, these tests include:

- 1- The IELTS test (General & Academic).
- 2- The English for Speakers of Other Languages (ESOL) tests.
- 3- The Business English Tests such as the Business English Certificate (BEC) and the Business Language Testing Service (BULATS).

### **2.3.2 The Educational Testing Service (ETS) tests in Saudi Arabia**

ETS in Saudi Arabia provides tests for candidates whose first language is not English and would like to study, train, live or work in a country where English is the first language.

ETS offers tests in Saudi Arabia mainly through the America-Mideast Educational and Training Services (AMIDEAST). For instance, these tests include:

- 1- Test of English as a Foreign Language (TOEFL) Paper Based Test (PBT)
- 2- TOEFL Internet Based Test (iBT)
- 3- TOEFL Institutional Testing Program (ITP)
- 4- TOEFL Junior
- 5- Test of English for International Communication (TOEIC)
- 6- TOEIC Bridge
- 7- Graduate Record Examination (GRE)

### **2.3.3 The Pearson English Tests in Saudi Arabia**

Pearson offers the Pearson Test of English (PTE), which is divided into three types:

- 1- PTE Academic (This is the only test offered in Saudi Arabia)
- 2- PTE General
- 3- PTE for Young Learners

### **2.3.4 The Graduate Management Admission Council (GMAC) Test**

The GMAC administers the Graduate Management Admission Test (GMAT), which is a test for postgraduate candidates who would like to do a business degree in higher education.

## **2.4 The College of Arts**

The College of Arts at King Faisal University (KFU) was opened in 2008. The College commenced its teaching programmes in the first semester of the academic year 2009/2010. Currently, it has four departments, namely: the Department of English Language (it was previously in the College of Education, Al Fraidan, 2011) where this study took place, the Department of Islamic Studies, the Arabic Language Department, and the Department of Social Sciences. The College is preparing to launch several new departments soon. These include the Department of Communication Information Technology, and Department of Libraries and Information (King Faisal University, 2012). The College offers undergraduate as well as postgraduate programmes. The College also gives scholarships for its staff who want to study abroad.

## **2.5 A brief description of IELTS candidates in Saudi Arabia**

This section gives a brief description of typical undergraduate and postgraduate Saudi students preparing or sitting for the IELTS test in Saudi Arabia and their writing experience. In this study, all of the students were preparing for or have taken the IELTS test. As mentioned earlier, English is taught in Saudi Arabia as a Foreign Language. Therefore, there is very little exposure to the English language outside the classroom. Yet, most Saudi students (i.e. undergraduate and postgraduate) have the basic command of English needed for communication.

Actually, some undergraduate and postgraduate courses at Saudi universities (e.g. Medicine, Computer Science, etc.) are taught in English. The students write their assignments, reports and research papers in English. In order to master these courses,

they must join a preparatory year programme in General English. The level of English differs according to the type of programme they want to study. All four skills are taught consistently. After completing the preparatory year programme successfully, students are taught English for specific purposes in their disciplines. Other students whose disciplines do not require a foundation year because they are not going to use English in their courses suffer from a lack of English language proficiency when applying to study at universities in English speaking countries. Therefore, most of them take English language courses (General then Academic) in language centres in Saudi Arabia or abroad in order to improve their level in English and compensate for this loss. In fact, the students in the Department of English are among the most fortunate ones because they are taught the English language skills from intermediate to advanced levels. Unlike other disciplines, the students receive intensive teaching in general as well as academic English. Therefore, they are expected to have a high command of English which is essential for successful completion of their content courses (e.g. English Literature, Applied Linguistics, etc.).

Despite their level in English, students who are seeking admission in English speaking countries need to show evidence to the receiving institutions that they have sufficient English language proficiency to survive at university. This is normally done through IELTS or TOEFL tests. With respect to the IELTS test, all Saudi students take the Academic Module because they are going to study at university. Most of these students enroll in IELTS preparation courses that are made up of intensive lessons focusing on developing language skills and test taking strategies needed to pass the test. The duration of these courses varies according to the level of English of those students.

Students can also prepare for the test through self-preparation booklets or materials from the Internet that have samples of the IELTS test. That is because they want to familiarise themselves with the test content and technique.

Whether the students prepared for the test by themselves or through language centres, this sort of preparation for the test helps the students to optimise their performance during the test. That is because these preparation courses or materials are teaching to the test. For example, the students are given samples of IELTS writing tests and they are asked to practice them for the real test. On the other hand, the writing courses that these students take at university are different as they focus on preparing the students to write research papers, reports, and long essays.

## **2.6 Conclusion**

This chapter has provided a general overview of the context of the study. It gave a detailed description of the educational system in Saudi Arabia. The current assessment practices taking place in Saudi Arabia were outlined, including the local and foreign testing and assessment organisations that offer tests for Saudi students. The chapter concluded with a brief description of the specific place of the study and the IELTS candidates in Saudi Arabia.

# CHAPTER 3

## **Chapter III**

### **Literature Review**

#### **3.1 Introduction**

This chapter reviews literature with respect to writing composition (process & product) and assessment, focusing on direct assessment of writing in second language contexts. Accordingly, it conceptualises the composition and the assessment of writing, in a socio-cognitive framework, as being both cognitive and social. More specifically, it investigates different theories of writing where writing is seen as a cognitive practice, a social construction practice, a social interaction activity, and as a socio-cultural practice. Therefore, candidates are placed in the centre of the assessment process rather than the test. This review sheds some light on the field of language testing and lists the types of tests with special reference to the IELTS test, which is the focus of the study. It then discusses validity, which is the focus of this study, as an essential quality in language tests.

In this review, emphasis is given to construct validity which is viewed by Messick (1989, 1996a, 1996b) as a unitary concept in which the three types of validity examined here, i.e. content validity, response validity and scoring validity, are seen as dependent on each other. That is to say, changes in the content of the test will affect the way candidates respond to it, and also when the scoring criteria are made known to candidates before the test they will have an impact on the processing of the task in terms of planning, monitoring and revision (Shaw and Weir, 2007). It should be noted that they are treated

separately throughout this study for descriptive purposes only. After discussing the theoretical construct, this review will shift into the practical side with special reference to the IELTS test. The focus will be on empirical studies on IELTS related to the types of validity being investigated in this study.

### **3.2 Theoretical framework**

Understanding the construction and scoring of texts used in the assessment of writing requires a theoretical framework that accounts for language use and assessment “as both a socially situated and a cognitively processed phenomenon” (Taylor, 2006, cited in Shaw and Weir, 2007, p.242). Therefore, the theoretical framework that underpins this study is informed by a socio-cognitive perspective that takes into consideration not only the cognitive processes that occur during academic writing but also the socio-cultural activities and interactional contexts in which the academic writing is carried out. This is necessary to account for the complex nature of academic writing in that it makes sense of the interactions that bind the writer, the process (task), the product (text), the reader and the context. Moreover, in adopting this multidimensional view, this study attempts to arrive at a comprehensive description of the above elements that may affect candidates’ performance in the IELTS Academic Writing Module (AWM). Taylor (2006) stressed that this may be best achieved through a socio-cognitive approach to writing (see Appendix III), as it:

helps promote a more ‘person-oriented’ than ‘instrument-oriented’ view of the testing/assessment process than earlier models/frameworks. It implies a strong focus on the language learner or test-taker as being at the centre of the assessment process, rather than the test or measurement instrument being the central focus (cited in Shaw and Weir, 2007, p.242).

In sum, this study adopts an integrated theoretical vision by arguing that writing is both cognitive and social (See Diagram 2). On the one hand, cognitivists from psychology have focused on the process of writing and highlighted the importance of the internal processes in text production; sociologists, anthropologists and sociolinguists, on the other hand, have paid attention to product and its role in the construction and reproduction of social relationships (Green, 2007a). It should be noted that this study assumes that there is a dialectical relationship rather than a dichotomous one between cognition and context. That is to say, they interact and affect each other where their boundaries are blurred. In explaining this dialectical interchange, Flower (1989, p.282) contended that “context cues cognition”. The following sections shed some light on each of these perspectives by drawing on relevant theories of writing. It should be noted that this study is not only concerned with writing and assessment but with writing assessment in L2.

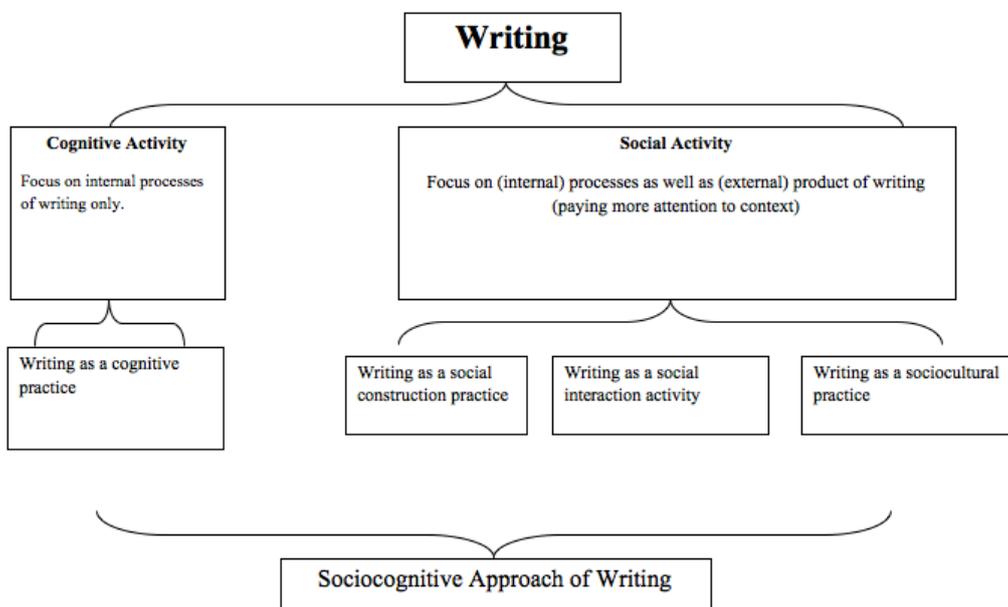


Diagram 2: Theoretical Framework of the Study

### **3.2.1 Writing as a cognitive practice**

Accounting for internal processes during writing lies at the heart of investigation for cognitivists. They see that an understanding of performance in writing can be obtained via descriptions of cognitive abilities involved during the writing process. In order to capture the various factors that distinguish individuals' performance and influence their mental activities during writing, models of L1 and L2 writing process have been proposed (Hayes and Flower, 1980; Bereiter and Scardamalia, 1987; Hayes, 1996; Grabe and Kaplan, 1996, Kellogg, 1996; Field, 2004). Also, these models are an attempt to explain the writing process through identifying the strategies that writers call upon during writing.

#### **3.2.1.1 Models of L1 writing**

In the following discussion of L1 models, this study will concentrate on two influential models of writing processes, namely, Hayes and Flower (1980) and that of Bereiter and Scardamalia (1987). Although these two models were developed based on studies on first language writing, they are worth mentioning due to their significant implications on second language writing research as “an L1 model of writing proficiency is commonly used as the metric in examining L2 writing” (Shaw and Weir, 2007, p.35).

##### **(a) Hayes and Flower Model (1980)**

This model was first introduced in 1980 as a reaction against popular models of writing which divided the composing process into three stages: the Pre-writing, Writing and Revision model (Rohman, 1965) or the Conception, Incubation and Production model (Britton, et. al., 1975). Hayes and Flower (1981, p.367) argued that “[t]he problem with

stage descriptions of writing is that they model the growth of the written product, not the inner process of the person producing it". In their model they proposed three key factors with respect to the act of writing: the task environment, the writer's long term memory and the writing process. The latter consists of three components, Planning, Translating and Reviewing, controlled by an internal monitor.

Despite the fact that it enhanced the way we think of the writing process, the model has been subject to a number of important critiques. Flower (1989, p.383) asserted that "[i]t failed to account for how the situation in which the writer operates might shape composing, and it had little to say about the specific conventions, schemata, or commonplaces that might inform the writer's 'long term memory' ". In addition, Berninger et al. (1996) listed a number of shortcomings with the model. Chiefly, they argued that the model "did not deal with individual differences" and it "did not make clear distinctions involving the *temporal dimensions* (before, during, or after translation) and *spatial dimensions* on which the planning and reviewing/revising processes operate (whole text or a portion of it)" (1996, p.197-8). Furthermore, the model does not provide a distinction between skilled and unskilled writing and "does not fully reflect the way in which writing varies according to task" (Shaw and Weir, 2007, p.35).

Above all, in the IELTS AWM test tasks, writing process approaches can be useful in understanding the composing process of writers, though they may not be applicable in the sense presented here due to the nature of the test (Cresswell, 2000). Because of the time constraints candidates may consider aspects like time management as top priority over a post-hoc revision or other processes. For this reason, timed tests of writing are scantily

suiting to the assessment of the process of writing as they assess the written product and thus give little evidence on how the text was composed (Hamp-Lyons and Kroll, 1997).

**(b) Bereiter and Scardamalia Model (1987)**

Bereiter and Scardamalia introduced this model in 1987 to address the shortcomings of Hayes and Flower's model. In their attempt to distinguish between novice and expert writers, they proposed two models: knowledge telling and knowledge transforming. The *knowledge telling model* can be best described as the generation of already known content knowledge to satisfy the requirements of a given task with little attention to planning or goal setting (Bereiter and Scardamalia, 1987). This process does not involve much effort and skill on the part of the writer. That is because recalling the appropriate kind of genre or finding the adequate response for a given task can be achieved with less elaborate problem-solving processes. Bereiter and Scardamalia (1987) have found that this model can explicate writing processes at both beginning and advanced levels.

The *knowledge transforming model*, however, embeds the former model as a sub-process and it is much more demanding in that it needs careful planning and goal setting in a problem-solving process that leads to the creation of new forms of knowledge (Bereiter and Scardamalia, 1987). It is through the process of writing that writers come to change their views and ideas as a result of development in their understandings of the texts with which they are engaged. In other words, the consequence of such an interaction with the text through problem-solving contributes not only to the development of text but also, most importantly, transformation of knowledge. Bereiter and Scardamalia (1987, p.12)

asserted that there exists “a two-way interaction between continuously developing knowledge and continuously developing text”.

While Bereiter and Scardamalia (1978) in their models have aided our understanding by making a useful distinction between skilled and unskilled writers, they have failed to give account of how writers progress from the knowledge telling model to the knowledge transforming model (Grabe and Kaplan, 1996). In addition, Green (2007a) noted that the social nature of writing was absent in these models. Furthermore, Susan Jones (2010, personal communication) contended that “[a] key feature of knowledge telling is the stringing of linked ideas, such that each idea triggers the next and so while a text might exhibit local cohesion there is not necessarily global cohesion”.

Despite these drawbacks, the models encompass some important issues in writing pedagogy and assessment. In particular, there are issues related to setting tasks which may also be germane to the IELTS AWM. Weigle (2002) illustrated this further, when referring to the models, by discussing the role of genre in determining the difficulty of the tasks. She noted that inexperienced writers can perform successfully in writing tasks of familiar genre through the use of knowledge telling processes, whereas even skilled writers may not perform well in writing tasks in an unfamiliar genre.

She found, though not empirically, that writing tasks demanding the use of knowledge telling “may not distinguish between better and poorer writers as well as tasks that are complex enough to elicit a knowledge transforming strategy from better writers” (Weigle, 2002, p.35). Clearly, such characterisation of writing tasks has resonance with different levels of performance in EAP tests. That is to say, low level tasks that have narrative or

instructional texts require knowledge telling skills while high level tasks that involve argumentative texts demand knowledge transforming skills (Shaw and Weir, 2007).

However, it should be noted that these models are interconnected as a result of the development process, as writers move from one model to another where the knowledge telling strategies are called upon when applying the knowledge transforming strategies. Green (2007a, p.35) rightly argued that “[t]he knowledge transforming model approach to writing does not exclude knowledge telling, but rather subsumes it”. Actually, the challenge for second language writers is to be able to transfer these skills which they have learned or acquired into second language contexts. This section below provides a supporting evidence of the composing strategies that second language learners transferred successfully into second language contexts.

### **3.2.1.2 Composing processes of second language writers**

Empirical research on writing processes of second language learners have focused on a number of variables characterising L2 writers’ performance (See Table 4 below).

Actually, most of these studies have found that writers’ characteristics affect the way they write in L2 (Zamel, 1983; Raimes, 1985; Cumming, 1989; Cohen, 1994; Spack, 1997; Boshier, 1998; Sasaki, 2000; Roca de Larios et al., 2008). Similarly, problem solving skills, as well as the quality of the text produced, have been found to be dependent on the level of proficiency in L2 and expertise in L1 writing.

Some of these studies have also shown that the higher the L2 proficiency and expertise in writing the more strategies are applied. For instance, skilled L1 and L2 writers allow

more time for planning before starting to write, generating ideas and revising whereas less skilled writers stop and translate more and revise words rather than larger sections of discourse during their writing due to insufficient time given for planning. Other studies have highlighted the impact of individual differences, socio-cultural and educational backgrounds, and context on the L2 composing process.

Furthermore, the use of L1 while composing in L2 has been shown to be sometimes useful but is dependent on aspects related to the context and the writer which include: the type of task undertaken (Wang & Wen, 2002), the topical knowledge related to L1 (Lay, 1982; Krapels, 1990; Friedlander, 1990) which are aspects of context and proficiency in the L2 and are one of the characteristics of the writer (Wang & Wen, 2002; Woodall, 2002; Wang, 2003; Wolfersberger, 2003). Likewise, proficiency in the L1 sometimes contributes to the quality of L2 texts. Weissberg (2000) suggested that L2 writers who are literate in their L1 benefited from L1 in the development of grammatical accuracy and most importantly in the emergence of new structures.

However, Van Weijen et al. (2009) found a negative correlation between L1 use during L2 writing and the quality of L2 text, where general writing proficiency also appeared to have a negative impact on L1 use during L2 writing. Other studies indicated that advanced writers made little use of their L1 while composing in L2 (Centeno-Corte's and Jime'nez Jime'nez, 2004; Beare and Bourdages, 2007).

In line with these findings, Silva (1993) examined 72 empirical studies that compared L1 and L2 in order to understand the nature of writing in L2. He found that there are significant differences between L1 and L2 writing in terms of “both composing processes (and sub-processes: planning, transcribing, and reviewing) and features of written texts (fluency, accuracy, quality, and structure, i.e., discoursal, morphosyntactic, and lexicosemantic)” (p.657). In fact, he confirmed this existing difference in his study as he reported that content generation and organisation in L2 were more difficult than in L1, goal setting and planning were less prominent, especially at the global and local levels, and that L2 writing was simpler in structure and style. Recently, Tillema et al. (2013) reviewed empirical studies which compared writing processes in L1 and L2 and pointed out that “the general conclusion is that L2 writing processes differ from L1 writing processes” (p.72).

Variable	Studies	Examples of Strategies	Studies
Expertise (skilled vs unskilled) in L1 and L2 writing, L2 proficiency and previous instruction	(cf. Sommers, 1980; Zamel, 1983; Raimes, 1985; Liebman-Kline, 1986; Jones & Tetroe, 1987; Moragne e Silva, 1988; Cumming, 1989; Cohen, 1994; Whalen & MeÂnard, 1995; Spack, 1997; Boshier, 1998; Sasaki, 2000; Roca de Larios <i>et al</i> , 2008; Van Weijden <i>et al</i> , 2009).		
Evidence of transfer of L1 writing strategies into L2 writing strategies have been documented by second language researchers	(Brooks, 1985; Jones & Tetroe, 1987; Cumming, 1989; Berman, 1994; Matsumoto, 1995; Beare, 2002; Wolfersberger, 2003, among others)	planning	(Jones & Tetroe, 1987; Uzawa & Cumming, 1989; Krapels, 1990; Woodall, 2002)
		translating	(Gerloff, 1987; Kobayashi & Rinnert, 1994; Uzawa, 1996)
		formulating	(Zimmermann, 2000; Krings, 1989)
		reading and writing extensively in L1 which could help in developing L2 competencies	(Brooks, 1985; Boshier, 1998)
		language switching	(Lay, 1982; Raimes, 1985; Jones & Tetroe, 1987; Cumming, 1989; Whalen & MeÂnard, 1995; Qi, 1998; Mancho´n <i>et al</i> , 2000; Zimmermann, 2000; Woodall, 2002)
		backtracking	(Raimes, 1987; Mancho´n <i>et al</i> , 2000; Wolfersberger, 2003)

		a- re-reading or re-scanning	(Raimes, 1987)
		b- back-translation	(Krings, 1987; Cumming, 1990; Whalen & Me'nard, 1995)
		“look-back” and the “look-ahead” principles	(de Beaugrande, 1984)
		solving problems and generating ideas	(Jones & Tetroe, 1987; Lay, 1982; Uzawa & Cumming, 1989; Krapels, 1990; Boshier, 1998; Roca de Larios <i>et al</i> 1999; Woodall, 2002; Wang, 2003).

Table 4: Summary of empirical research on writing processes of second language learners

### 3.2.2 Writing as a social construction practice

Writing during the 1970s witnessed an emphasis on the composing process while in the 1980s it undertook a shift in focus towards the importance of social context in the composing process. Unlike the traditional cognitive view of writing that focused merely on internal processes or that theoretical perspective on writing from linguistics which considered the textual features of the product, the current social view of writing takes into account external factors or the social context influencing performance in which such processes take place. In doing so, the act of writing becomes social as it is carried out in a social context.

Hamp-Lyons and Kroll (1997, p.8) contended that writing is “an act that takes place within a context, that accomplishes a particular purpose, and that is appropriately shaped for its intended audience”. Likewise, Hyland (2002, p.69) asserted that “all acts of writing are part of wider social and discursive practices which assume certain things about relationships between those involved and how practices are carried out”. Therefore, a text should be viewed as a joint product of an individual and a certain context. In illustrating the social dimension of writing, Hayes (1996, p.5) maintains that “[w]hat we write, how we write, and who we write to is shaped by social convention and by our history of social interaction”.

In second language contexts where academic writing is used, “most writing for academic classes is in response to a specific assignment or prompt” (Johns, 1986, p.253). Very often, students are introduced into a ‘discourse community’ (Spack, 1988). In such a community they are socialised into the discourse conventions of a particular discipline. Gaining access to such a community through writing requires not only a command of grammar and vocabulary but also an examination of “the kinds of issues a discipline considers important, why certain methods of inquiry and not others are sanctioned, how the conventions of a discipline shape text in that discipline, how individual writers represent themselves in a text, how texts are read and disseminated within the discipline, and how one text influences subsequent texts” (Spack, 1988, p.38).

In this way, writers obtain membership and create identity through discourse which in turn let them “locate themselves and their own ideas with respect to other ideas and texts within their communities” (Hyland, 2002, p.69). He added that writing then becomes a “socially situated response to particular writing contexts and communities” (Hyland,

2003, p.17). Academic writing, then, becomes a single appreciated set of practices which are socially contextualised rather than as one universal set of cognitive skills (Grabe and Kaplan, 1996). This implies that writing in general has a wide range of functions, not all of which are accounted for in the academic context.

In the context of IELTS, for instance, candidates' writing is a result of complex social interactions shaped by the context. Their texts are a response to the tasks' prompts which are designed by the test developer in order for the candidates to meet the expectations of the markers and the specifications of the test itself. It can be seen here that the interest is in the production of texts for the sake of assessment rather than for communicating information. Sullivan (1997, p.55) argued that "[i]t is not too much to say, then, that high stakes assessment, such as placement testing, operates as a site of conflict, on which students' institutional identities as writers are defined and contested". Obviously, this could be because under test conditions the content and the style of response have already been determined by the test designer or developer. Starfield (2004, p.68) pointed out that "[i]n this type of writing, the roles of writers and readers are shaped by the constraints of the contexts: the mode of discourse is not only written to be read, it is written to be assessed".

To sum up, it has become clear that the social context has an effect on the kind of text or genre produced and that writers need to be aware of that when composing their texts in order to meet the demands of that context. The context also informs the value and purpose of the text and the social positioning of the writer. Based on this premise, writing involves writers in an interaction process with their target audience or readers.

### 3.2.3 Writing as a social interaction activity

According to Nystrand (1989) writers interact with readers in the sense that “[t]he process of writing is a matter of elaborating text in accord with what the writer can reasonably assume that the reader knows and expects, and the process of reading is a matter of predicting text in accord with what the reader assumes about the writer’s purpose” (p.75). That is to say, the writer as well as the reader adopts a reciprocal role that is based on the assumption that both of them should be capable of interpreting what is written or read. Making sense of the written text is critical to the extent that interaction is highly unlikely to take place without it. In short, texts bridge the interests of writers and readers and, most importantly, they are not just a result of the writing process but also a means of communication between both parties (Nystrand, 1989).

In a call to adopt a social perspective of interaction rather than the cognitive one that dominated most studies in language testing until recently, McNamara (1997, p.459) recommended that there should be a “renewed focus on the social dimension of interaction”. Participation in the social nature of interaction requires a prior knowledge of the communicative conventions (e.g. how to match the form of language to the topic, the setting, the interlocutor, and so on) that can be part of the cognitive dimensions distinguishing successful candidates (McNamara, 2000). The interaction with the social context is reflected in the test design in terms of handling the *test content* (based on a socio-linguistically inspired needs analysis of roles, settings, participants and so on in the target language use situation) and engagement with test materials in *test process* (represented by the operation of *strategic competence*, by which the candidate mentally

interacts with features of the *situation of target language use* such as planning and judgement, and the like) (McNamara, 1997, p.450, his italics).

Similarly, in using texts as a medium of communication, successful writers always think about their audience or readers when they write. Ede and Lunsford (1984) have proposed two types of audience: audience addressed and audience invoked. Audience addressed is the real already defined audience whose attitudes, beliefs, and expectations are essentially known to the writer. On the other hand, audience invoked refers to an imagined audience that is constructed by the writer for rhetorical purposes. Ede and Lunsford (1984, p.160) pointed out that the main role of the writer in audience invoked is not “to analyze an audience and adapt discourse to meet its needs” as in audience addressed but rather to use “the semantic and syntactic resources of language to provide cues for the reader - cues which help to define the role or roles the writer wishes the reader to adopt in responding to the text”. However, they argue that the audience invoked position “distorts the processes of writing and reading by overemphasizing the power of the writer and undervaluing that of the reader” (Ede and Lunsford, 1984, p.165).

Clearly, adopting either role indicates that the audience has a significant influence on the production of discourse. More specifically, Grabe and Kaplan (1996, p.207–208) have identified five parameters that have an impact on the decisions made by the writer during the process of text production. These include:

- 1- The number of persons who are expected to read the text.
- 2- The extent of familiarity of the readers with the writer.
- 3- The status of the readers in relation to the writer.

4- The extent of shared background knowledge between the readers and the writer.

5- The extent of shared specific topical knowledge between the readers and the writer.

In line with the audience invoked model, both of IELTS AWM tasks have an imagined audience. For instance, in the first task the addressee is a university lecturer and in the second task it is an educated reader with no specialist knowledge (Green, 2007a). In contrast to other researchers (see Douglas, 2000) who call for the specification of “audience” in the task rubrics, Green (2007a, p.57) argued that this may cause confusion on the part of the candidate as “[i]t is not made clear whether the tasks are to be evaluated in the manner of the imaginary ‘university lecturer’ and ‘educated reader’ to whom the text is supposedly addressed...or in the manner of an English language teacher, perhaps with opposing priorities”. In other words, specifying the audience in the task prompts may present a challenge for the candidates in applying the appropriate criteria that meet the expectations of the examiner.

### **3.2.4 Writing as a socio-cultural practice**

An important issue that is ignored in the cognitive paradigm according to Prior (2006, p. 54), is that it is “too narrow in its understanding of context and was eclipsed by studies that attended to social, historical, and political contexts of writing”. The socio-cultural theory (SCT) of mind introduced by Vygotsky, a Soviet psychologist, is a theory that takes into account the role of social and cultural interactions in developing cognition. In order to account for learning and development it views writing as a process rather than a product (Lantolf and Thorne, 2006). In integrating mind and body into a single unity, it begins with the assumption that actions are inseparable from the social contexts in which

they occur (Wertsch, 1991) and that consciousness is socially situated in interactions rather than residing inside the head of an individual (Vygotsky, 1978).

Another key assumption in the SCT which is relevant to this study is mediation.

According to Lantolf (2000, p.1), Vygotsky argued that “just as humans do not act directly on the physical world but rely, instead, on tools and labour activity, which allows us to change the world, and with it, the circumstances under which we live in the world, we also use symbolic tools, or signs, to mediate and regulate our relationships with others and with ourselves and thus change the nature of these relationships”. To put it differently, mediation is the use of tools or signs for the sake of changing the nature of our relationships with the individuals or the world we live in. In the IELTS test, for instance, candidates interact with texts by using language as a tool which mediates this interaction. Poehner (2007, p.326) stressed that “individuals are always mediated by cultural artefacts, social practices, and activities” and that “[t]hey are mediated even when they are working alone, in which case their cognitive functioning is mediated by their history of interactions with the world”. In this way, candidates under test conditions bring their past experience of something or background knowledge of a topic and their linguistic knowledge to mediate their interaction with texts in order to make meaning and carry out the task at hand.

Brooks (2009, p.361) supports this view and argues that “individuals bring with them their agencies, histories, and personalities to a test situation so if anything is predictable, it is that performance will vary”. Obviously, the mediated interaction here is between the language user (his linguistic and background knowledge) the context and the discourse (Bachman, 1990).

A further important assumption in the SCT is internalisation. As viewed from an SCT perspective, internalisation is the “process through which a person moves from carrying out concrete actions in conjunction with the assistance of material artefacts and of other individuals to carrying out actions mentally without any apparent external assistance” (Lantolf, 2000, p.14). This indicates that SCT assumes that all cognitive processes are constructed through social interactions with individuals, artefacts or tools in the environment rather than pre-existing inside the mind waiting to emerge at a specific moment of an individual’s lifetime. Indeed, the development of cognitive abilities moves from dependence on external assistance into relying on internal assistance for mediation. When an individual has internalised certain aspects of social interactions, he/she is able to “self-mediate, or self-regulate to use Vygotsky’s term” (Poehner, 2007, p.326). That is to say, the individual relies mainly on internal support for mediation. Internalisation in this sense may have links to McLaughlin’s (1990) concept of automaticity in which individuals gain control over linguistic knowledge through repeated encounters with skills to be learned. The difference is, however, that individuals can be automated through practice (the more these skills are routinised the easier they can be applied), unlike internalisation where intervention in the form of mediation enables individuals to perform a task alone with no external assistance. In other words, automaticity aims at putting something to use efficiently, whereas internalisation seeks to foster independence in individuals by freeing them from external support.

In the case of the IELTS AWM test, candidates depend mainly on themselves in order to carry out the test tasks. That is because decisions on whether they have adequate proficiency to perform at a certain level are largely based on how successful they are in

composing appropriate texts independently within the allotted time. In this way, they no longer rely on external assistance (which is actually prohibited as it is considered to be a form of cheating) but rather depend on their internal capacity. They 'self-mediate' their interactions with the test tasks as they retrieve internalised external social interactions. For instance, most candidates who come to the test prepared may have received assistance in the preparation through preparation books, preparation courses, internet websites, friends or themselves as repeaters of the test.

Successful candidates are those who are able to internalise these external forms of preparation for the test so that they can accomplish the test tasks alone with no intervention for assistance. For a better understanding of their performance, inferences need to be based on the assumption that activity is inseparable from the context in which it is carried out.

### **3.3 A brief introduction to language testing**

Allen (2009) defined language testing as “the practice and study of evaluating the proficiency of an individual in using a particular language effectively”. Since the middle of the last century, language testing has responded to shifts in paradigms in language teaching and learning. Testing practices were characterised in the Post War era by being subjective, where traditional techniques in testing like dictation, essay, composition and translation were frequently applied. This impressionistic view of testing was replaced by scientific (objective) theories based on structuralism and behaviourism. Successful language learners were described as being able to recognise different sounds, grammatical structures and vocabulary items. These two theories were soon replaced by audiolingualism. Repetition drills were at the heart of this method with a focus on stress, intonation and rhythm.

Currently, communicative methods are stressing the importance of assessing language use rather than language form. Canale and Swain (1980) were the first pioneers to suggest a communicative competence model based on Hymes’ (1966) theory of communicative competence in which he was the first to coin these terms together. It can be noted that “language tests are under somewhat of a challenge, as they respond to critiques of individualistic notions of performance and are increasingly being scrutinized for their social accountability, in line with the critical turn in applied linguistics generally” (McNamara, 2006, p.763). More specifically, Skehan (1991, p.4) pointed out that “[t]esters expect input from linguists for models of language and performance, from psycholinguists and second language acquisition researchers for insight on processes of

language use and learning, and from sociolinguists for perspectives on such issues as language variation and the relationship between language and context". Fortunately, having received such large inputs from other fields, language testing now has a central place in applied linguistics (Davies, 1991).

### **3.3.1 Types of tests**

The process of test construction is affected by the purpose that test developers have in mind before or during the stages of test design. Accordingly, this predefined purpose will determine the type of test produced. That is to say, different types of tests will have different purposes. The overriding purpose for most tests is to pinpoint the areas of strengths and weaknesses in candidates' performance. Similarly, evaluating the efficiency of a certain programme is another important purpose in test construction. These two purposes have broad applications in the literature through test types. It seems that there is no agreed number of test types in the literature.

The most commonly used typology of tests is that of Alderson et al. (1995), where they listed a simple and straightforward typology of tests which differed according to purpose. These types included placement tests, progress tests, achievement tests, proficiency tests and diagnostic tests. The following section gives a brief summary of these types.

#### **3.3.1.1 Placement tests**

The aim of these tests is to determine the student's level of knowledge so that they can be placed in the appropriate course or class where the content of the test is based on a syllabus or unrelated materials (Alderson et al., 1995). In the case of IELTS, where the

content of the test has no relevance to previous materials, students' overall band scores can be used for placement purposes at a certain level in a language centre.

### **3.3.1.2 Progress tests**

These tests are designed to check if the students are making progress in a course or a programme by fulfilling predefined aims and objectives. A common characteristic of these tests is that they are spread throughout the course or programme to reflect on students' knowledge of certain taught materials.

### **3.3.1.3 Achievement tests**

Achievement tests have something in common with progress tests as they measure the amount of learning that students have as a result of instruction. Yet, achievement tests tend to be longer and more comprehensive as they are given at the middle or at the end of the course. Consequently, they cover more content and materials as they are based on more objectives.

### **3.3.1.4 Proficiency tests**

The purpose of constructing proficiency tests is to measure the students' language capacity by describing the extent to which an individual is able to survive in the target situation. This description shows the amount of learning acquired till the date of assessment which can be either in a general or specific domain of knowledge. The content of the test is not based on a specific course of instruction in a language programme. The IELTS test is an example of a proficiency test which has a general and

an academic module in which candidates are enrolled according to the kind of training or study required in the target situation.

### **3.3.1.5 Diagnostic tests**

Diagnostic tests refer to tests that are designed for the intention of pinpointing the areas of weakness where remedial work is required. Such tests have the potential of providing us with informed judgments on a particular skill or sub-skill that needs further development.

To sum up this discussion about test types, it is noteworthy to point out that while this study focuses on the IELTS test it deals with it as being both a placement test and proficiency test. This consideration is built on the assumption that some participants who come from language centres may have taken the test at some stage during the language course. That is to say, their level has been determined via the test scores by placing them at the level corresponding to their current competence and performance. On the other hand, other students took the test in order to gain direct entry to university because they have met the requirements of that university or college. In this case, it is a proficiency test where their scores in the test show that they are not in need of further language development as they have a full command of language usage in the specific domains needed in the target situation. As a result, this study argues that it is not only the purpose of the test that determines the kind of test but also the test scores that candidates obtain in the test. In other words, a test which is designed to be a placement test can be used as a proficiency test and vice versa depending on the test scores.

### **3.4 Essential qualities in language tests**

Validity and Reliability are two essential qualities in any test because they “provide the major justification for using test scores—numbers—as a basis for making inferences or decisions” (Bachman and Palmer, 1996, p.19). The focus of this study is on validity because it gives meaning to our test rather than form (Davies, 2012). Moreover, Hamp-Lyons added that “[n]o test can be valid without first being reliable” (1991a, p.52). In other words, we can have reliable test which is not valid but we cannot have a valid test which is not reliable. This shows that it is validity which is central in assessment practices and what tests may need most. That is why it is investigated in this study from the perspectives of the stakeholders.

#### **3.4.1 Validity**

The most explicit definition of validity that most researchers are likely to agree on is the one by Garrett (1937, p.324) where he defined validity as “the extent to which the test measures what it purports to measure”. The key word in this definition is “purports” which refers to the purpose or intention for which the test is designed. Carmines and Zeller (1982, p.17) argued that “one validates not the measuring instrument itself but the measuring instrument in relation to the purpose for which it is being used”. Therefore, a valid test could be valid for one purpose but invalid for another. Similarly, in an attempt to modify the general view of validity, Weir (2005, p.12) argued that “validity resides in the scores on a particular administration of a test rather than in the test *per se*”. Validity, in this sense, is a feature of the inferences made on the basis of test scores and use rather than a characteristic of the test itself (Alderson and Banerjee, 2002). Traditionally, building a validity argument used to be based on statistical analyses only but at present

(which is what this study is following) the students' perceptions about their test taking behaviours and what the test is measuring are being asked in current studies about test validity (Cohen, 2012).

Validity has always been an essential aspect of any measurement tool. Davies (1991) argues that tests have three demands which must be equally attended to in a language test: the language (has to do with validity), the learning (has to do with washback) and the measurement (has to do with reliability). Validation, as Weir (2005, p.15) puts it, "can be seen as a form of evaluation where a variety of quantitative and qualitative methodologies are used to generate evidence to support inferences from test scores".

A major recommendation by the ELTS Validation Report for future research on test development was that validation should be conducted before the new test is administered to candidates (ELTS test did not have a validation study before it was put into operation) (Cripper and Davies, 1988). This means that the validation of tests especially for those used for the assessment of language proficiency should be an on-going practice, particularly as tests are triggered by revisions that are carried out on their content, construct or administration through the prologue of new tasks or materials. As these revisions are undertaken and their recommendations implemented, they create new challenges to candidates who are going to sit the test. Therefore, it becomes a necessity to conduct validation studies on the test, especially when it is a high stakes test where the results will affect the future of candidates.

Admittedly, this underlines the importance of the test making strong claims to validity to ensure that candidates' performance is measured by fair standards. It should be noted,

however, that the level of validity varies according to the purpose of the test. Most importantly, perfect validity cannot be achieved. That is to say, there is no 100% accurate measure of the candidates' performance in any test. Bachman (2010) argued that it is a misconception to think that there is a best test and also terms such as good or bad are not helpful in describing the qualities of a language test. Nonetheless, one needs to construct or look for the most accurate measures that are likely to reflect or predict the actual performance in order to minimise the risk of producing less valid inferences on test scores and use.

#### **3.4.1.1 Types of validity**

Validity is a broad term that can cover a range of issues in language proficiency tests. The types of validity below show clearly how far validity can contribute to the integrity of a test. There is variation in the classification as well as the number of types of validity in the literature. Campbell (1957) evaluated validity in terms of two criteria, i.e. internal and external (he referred to external validity as “generalizability” or “representativeness”). These two types were later extended into four: statistical conclusion validity, internal validity, construct validity and external validity (Cook and Campbell, 1979). Carmines and Zeller (1982) have described three types of validity: criterion-related validity, content validity and construct validity. However, the revised Standards for Educational and Psychological Testing regard criterion-related, content and construct validity as types of evidence rather than types of validity.

Wall et al. (1999) have identified six types of validity, i.e. content validity, predictive validity, concurrent validity, construct validity, response validity, and washback validity. Brualdi (1999) made a distinction between traditional (i.e. criterion-related, content and

construct validity) and modern concepts of validity, e.g. Messick (1989, 1996a, 1996b) who defined six aspects of validity: content, substantive, structure, generalizability, external factors and consequential aspects. Unlike traditional concepts of validity which are incomplete and fragmented, Messick, in his modern approach, viewed validity as a unified concept that tells us how the test is used (Messick, 1989, 1996a, 1996b). Weir (2005) listed five types of validity in a framework to conceptualise writing test performance: context validity (traditionally known as *content validity*), cognitive validity (this is very similar to the traditionally known *response validity*), scoring validity, consequential validity and criterion-related validity. Actually, the most useful detailed distinction or classification of the types of validity according to research paradigms is the study by Onwuegbuzie and Johnson (2006). They classified types of validity according to research paradigm: quantitative, qualitative and mixed methods research.

It is noteworthy to mention their discussion about qualitative researchers who tend to replace traditional quantitative terms about types of validity with more acceptable qualitative ones. Onwuegbuzie and Johnson (2006) considered Lincoln and Guba (1985) as an example when they replaced credibility for internal validity, transferability for external validity, dependability for reliability and confirmability for objectivity.

Moreover, Onwuegbuzie and Johnson (2006, p.51) argued that “validity has generally been replaced by the term trustworthiness within qualitative research”. In mixed methods research they listed nine types of validity which they re-named legitimations instead of validity: sample integration legitimation, inside out legitimation, weakness minimization legitimation, sequential legitimation, conversion legitimation, paradigmatic mixing

legitimation, commensurability legitimation, multiple validities legitimation and political legitimation (Onwuegbuzie and Johnson, 2006).

However, I would argue that this list, i.e. nine types of validity in mixed methods research, may not attend to the specific level of interpretation about validity on which other approaches try to focus. Nevertheless, the aim in listing these types of validity in mixed methods research seems to be about giving equal attention to presenting quantitative and qualitative approaches. In other words, validity in mixed methods research ensures that there is a successful mixture between quantitative and qualitative approaches. It is important to know that, among the types of validity mentioned in this review, only three types of validity i.e. criterion-related, content and construct validity, are most commonly referred to in the literature.

In this study, I will discuss in more detail three types of validity, i.e. content validity, response validity and scoring validity, as they are related to the focus of investigation and questions of my study. Moreover, these types of validity are dependent on each other. In other words, changes on the content of the test will affect the way candidates respond to it and also when the scoring criteria are made known to candidates before the test they will have an impact on the processing of the task in terms of planning, monitoring and revision (Shaw and Weir, 2007). Accordingly, the IELTS test will be examined from the perspective of each of these three types of validity and will thus give different answers to the research questions. It should be noted, however, that this study considers these three types of validity as constituting a broader and unified concept of test validity which is referred to as construct validity (Messick, 1989). Therefore, they are treated separately throughout this study for descriptive purposes only.

### **3.4.1.1.1 Content validity**

Hatch and Farhady (1982, p.251) defined content validity as “the extent to which a test measures a representative sample of the subject matter content”. The key word in this definition is “representative” which indicates that the test should include items and tasks that students are likely to encounter in the “target situation” (Wall et al., 1999). In the case of the IELTS test as a language proficiency test, the test has to achieve the following aspects in order to be valid in content. These aspects are content relevance and content coverage (Bachman, 1990).

According to Messick (1980, p.1017), content relevance refers to the “specification of the behavioural domain in question and the attendant specification of the task or test domain” (cited in Bachman, 1990, p.244). The second important aspect in investigating content validity is content coverage. Bachman (1990, p.245) defined content coverage as “the extent to which the tasks required in the test adequately represent the behavioural domain in question”. Undoubtedly, test specifications that tell us about test content and relevance are crucial to content validity as they can inform test takers and developers about the level of representativeness of the test.

The two types of validity mentioned above have been labelled differently. Nevertheless, such differences in labelling still approximate the aforementioned types of validity, bearing the same connotations. These elements, as the researchers prefer to call them, include: transparency (clear, accurate information about the test, such as outcomes to be evaluated, formats used, weighting of items and sections, time allowed to complete the test, and grading criteria [Lloyd et al., 2005]) and usefulness (designing tests with a

specific purpose in mind, a defined group of candidates and a particular language use [Bachman and Palmer, 1996]). It can be seen that these two elements demonstrate that test content is a necessary aspect of test use and development, where the aim is to produce test content that is comprehensive in terms of coverage and relevant to what the examinees will do in the target situation. In essence, transparency and usefulness are just other ways of approaching content validity through coverage and relevance.

However, Geisinger (1992, p.207) claimed that in those types of content validity neither test scores nor test responses are considered “[h]ence, numerous test theorists (e.g. Guion, 1977; Tenopyr, 1977 among others), have suggested that what has traditionally been called content validity is not truly validity at all”. In order to overcome this problem of content validity this study has set out to investigate response validity with a focus on test scores and responses to support the evidence from content validity.

### **-Authenticity**

The first appearance of this term in language testing literature was in the 1970s, having emerged earlier from applied linguistics in the mid 1960s at a time when communicative methodology was gaining popularity in teaching and testing real life language (Lewkowicz, 2000). Authenticity is regarded as a crucial aspect of language tests (cf. Lynch, 1982; Morrow, 1991; Wood, 1993; Bachman and Palmer, 1996; Douglas, 1997). It is defined by Bachman and Palmer (1996, p.23) as “[t]he degree of correspondence of the characteristics of a given language test task to the features of a TLU task”. According to this definition, a test is said to be authentic if its content matches that of the Target Language Use (TLU) domain. This, however, does not necessarily indicate that there

exists a dichotomy of authentic and inauthentic tasks as the concept of authenticity is rather a complex one (Breen, 1985). Therefore, he suggested (together with other researchers cf. Bachman, 1990; 1991; Anderson et al., 1996) that authenticity is a relative feature of a test. In his view, authenticity is reported on a continuum so that the more the task in the test is related to real life performance the higher the authenticity and vice versa.

Yet, it was not clear how the degree of correspondence (high or low) between test tasks and TLU would be estimated until Bachman and Palmer (1996) proposed a framework of task characteristics (see Table 5) that was built on an earlier one proposed by Bachman (1990). In this groundbreaking framework which is useful in determining the authenticity of tasks, designing test tasks and assessing reliability, they listed five elements which matched tasks in both test and non-test domains with respect to setting, test rubric, input, expected response and the relationship between input and response. This framework is comprehensive because it takes into account the input and the expected response as well what test takers are likely to produce as a result of their interactions with the characteristics of the test tasks. However, this framework has failed to point out how to identify the critical features of TLU tasks (Lewkowicz, 2000). She added that if characterisation of TLU tasks were possible then “[t]he large number of TLU tasks characterized could ensure a level of authenticity for most test tasks selected, since the larger the number of TLU tasks to choose from, the more likely it is that there would be a level of correspondence between the test tasks and the TLU domain” (p.51).

<b>Task Characteristics</b>				
Characteristics of the setting	Characteristics of the test rubrics	Characteristics of the input	Characteristics of the expected response	Relationship between input and response
<p><i>Physical characteristics</i></p> <p><i>Participants</i></p> <p><i>Time of task</i></p>	<p><i>Instructions</i></p> <p>Language (native, target) Channel (Aural, visual) Specification of procedures and tasks</p> <p><i>Structure</i></p> <p>Number of parts/tasks Salience of parts/tasks Sequence of parts/tasks Relative importance of parts/tasks Number of tasks/items per part</p> <p><i>Time allotment</i></p> <p><i>Scoring method</i></p> <p>Criteria for correctness Procedures for scoring the response Explicitness of criteria and procedures</p>	<p><i>Formal</i></p> <p>Channel (aural, visual) Form (Language, non-language, both) Language (native, target, both) Length Type (item, prompt) Degree of speediness Vehicle ('live', 'reproduced', both)</p> <p><i>Language of input</i></p> <p><b>Language characteristics</b></p> <p>Organisational characteristics Grammatical (vocabulary, syntax, phonology, graphology) Textual (cohesion, rhetorical/conversational organisation)</p> <p><b>Pragmatic characteristics</b></p> <p>Functional (ideational, manipulative, heuristic, imaginative) Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)</p> <p><b>Topical characteristics</b></p>	<p><i>Formal</i></p> <p>Channel (aural, visual) Form (language, non-language, both) Language (native, target, both) Length Type (selected, limited production, extended production) Degree of speediness</p> <p><i>Language of expected response</i></p> <p><b>Language characteristics</b></p> <p>Organisational characteristics Grammatical (vocabulary, syntax, phonology, graphology) Textual (cohesion, rhetorical/conversational organisation)</p> <p><b>Pragmatic characteristics</b></p> <p>Functional (ideational, manipulative, heuristic, imaginative) Sociolinguistic (dialect/variety, register, naturalness, cultural references and figurative language)</p> <p><b>Topical characteristics</b></p>	<p><i>Reactivity</i> (reciprocal, non-reciprocal, adaptive)</p> <p><i>Scope of relationship</i> (broad, narrow)</p> <p><i>Directness of relationship</i> (direct, indirect)</p>

Table 5: Task Characteristics (Adapted from Bachman and Palmer, 1996, p.49-50).

Nonetheless, the collection of materials from the TLU is not always easy and straightforward. That is to say, the more general the tasks in the TLU to be included in the test the easier the selection would be for language test developers, whereas the more specific the tasks the more difficult it would be to find appropriate tasks that mirror the authentic TLU domain, and this may be the reason why test developers give little attention to authenticity when compared with other aspects such as reliability (Weigle, 2002). She further commented that authenticity has another problematic area with respect to testing English for academic purposes (especially writing). She argued that a timed writing test with a topic unknown to students in advance is “relatively inauthentic in at least four ways” (p.52). These are as follows:

- 1- In academic writing contexts (e.g. at university), students have the advantage of knowing the topic and dealing with it through reading, speaking and listening before writing and therefore have background knowledge and schemata within which to write.
- 2- Most academic writing is not timed or speeded and students can make use of outside resources, and revise and edit their work before submission.
- 3- The audience for writing tests is frequently unknown to the candidates while in most academic writing contexts the audience is the instructor and that gives them an idea of how their texts should look like in order to meet the expectations of the marker.
- 4- The primary judgment in scoring criteria for most academic writing is on accuracy of content, rather than on the appropriateness of the organisation or the use of language (Weigle, 2002).

Other researchers, like Doye (1991), seem to join the debate in problematising authenticity by arguing that a complete congruence of test situation and real-life situations is impossible because “a language test is a social event that has - as one of its characteristics – the intention to examine the competence of language learners” (p.5). He quoted Pickett (no date) when asserting that “[b]y virtue of being a test, it is a special and formalised event distanced from real life and structured for a particular purpose. By definition it cannot be the real life it is probing” (p.5). Similarly, Spence-Brown (2001, p. 464) noted that “[t]he basic problem for language testing is that tests are, by their very nature, artificial contexts for language use”. That is to say, candidates in testing situations are prompted to display their knowledge and skills and thus the test interaction is said to follow different rules from that of the non-test interaction (Spolsky, 1985).

This may be why it is difficult to have completely authentic tests that correspond to non-testing contexts. Alan Davies (2008, p.69), a prominent scholar in language testing who has long been involved in developing and validating language tests, shares this view as he rightly contends that “tests simply cannot be authentic: what they can (all they can) do is to simulate authenticity”. He pointed out that this is the position that the IELTS test adopted ultimately during the mid 1990s in the second revision. Based on this premise, the IELTS tasks do not claim authenticity in the strong sense of the definition but a simulation of those tasks that test takers would be expected to encounter or perform in the TLU domain outside the testing contexts. This understanding of authenticity is in line with the theoretical developments in language teaching and learning in what has been known as communicative language testing in which the focus of testing has shifted from

testing behaviours to testing abilities (Davies, 2008). That is to say, the aim of communicative language tests, e.g. IELTS, is to test language use rather than language knowledge. Luoma (2001) rightly described that “[j]udging by the test format, the developers of IELTS supported a communicative or, more precisely, an interactionist view of language with an emphasis on actual language use” which means that “it was considered important to assess writing through productive writing tasks” (p.233-234).

In sum, advocates of authenticity call for test tasks to be taken from real life contexts while those who are doubtful of such practice argue that what tests can best do is to simulate authenticity. For language testers, moreover, generalisations across contexts is the ultimate aim of establishing authenticity in any language test but this may not be the case for other stakeholders, namely test takers involved in the testing process for whom authenticity seems not to be so important (Lewkowicz, 2000).

#### **3.4.1.1.2 Response validity**

According to Henning (1987, p.96), response validity is “the extent to which examinees responded in the manner expected by the test developers”. Shaw and Weir (2007) referred to cognitive validity to mean response validity. They considered it in a writing task as “a measure of how closely it represents the cognitive processing involved in writing contexts beyond the test itself i.e. in performing the task in real life” (ibid, p.34). They identified six cognitive validity parameters in writing (see Figure 1). Due to the significance of these parameters involved in the cognitive validity in writing, this study will discuss them in the following sections.

<b>Cognitive Validity</b>
<p><b>Cognitive Processes</b></p> <ul style="list-style-type: none"> <li>• Macro-planning</li> <li>• Organisation</li> <li>• Micro-planning</li> <li>• Translation</li> <li>• Monitoring</li> <li>• Revising</li> </ul>

Figure 1: Cognitive validity parameters in writing (Adapted from Shaw and Weir, 2007, p.34)

**(a) Macro-planning**

In this aspect of processing writers are involved in ideas generation in order to fulfil the requirements of a task by depending on their content knowledge (Field, 2004). They set the goal for the task which “includes consideration of target readership of the genre of the text (earlier experience as a reader may assist) and of style (level of formality)” (ibid, p.329).

**(b) Organisation**

Ideas which have been generated already are provisionally organised in abstract form in relation to the text as a whole as well as in relation to each other (Field, 2004). Writers evaluate the ideas according to their “relative importance, and decisions made as to their relative prominence in the text” (ibid, p.329). He added that this process may result in the form of rough notes by the writer.

### **(c) Micro-planning**

Writers in this stage of processing start to plan at sentence and paragraph level, with constant reference back to decisions already made in previous stages and to the way in which the text has developed so far (Field, 2004). He explained that “[a]t this stage, the writer needs to give consideration to whether an individual piece of information is or is not shared with the reader (a) by virtue of shared world knowledge or (b) as a result of earlier mention in the text” (ibid, p.329).

### **(d) Translation**

In this processing writers convert the assembled text produced so far from abstract to linguistic form during which the exact language of the text vary from writer to writer and from task to task (Field, 2004). He added that some writers may be able to construct a full sentence at this stage while others may store the information they have in the form of key words in a writing buffer which appears to be phonological in form (ibid). This cognitive processing has a number of explanations. Field (2004) explained that “[i]t may be that a spoken trace is more robust than a written one. Or it may be that future content stored in phonological form is less likely to interfere with the process of putting words on to the page in visual form” (p.329-30).

### **(e) Monitoring**

Writers during this processing of self-monitoring undergo a complex operation at many different levels of representation. Field (2004) described that writers at the lowest levels are concerned with checking the accuracy of spelling of spelling, punctuation and syntax whereas at higher levels writers examine the text produced so far to check its relevance to

their goals and to the overall development of the discourse structure of the text.

Monitoring is a cognitively demanding activity and therefore writers are likely to focus on one level of analysis at a time where attention, while producing a text, is given “to lower-level features, with higher-level ones reserved for a later editing stage” (ibid, p.330).

#### **(f) Revising**

In this aspect of processing writers return to features of the text developed so far in order to revise it. This process is recursive rather than linear (Hayes and Flower, 1981) and takes place at the level of sentence, paragraph or the whole text. Field (2004) argued that “[m]onitoring, editing and revising are thus feed-back processes which can impact upon any of the previous stages, causing the writer to revise a macro-plan, to reword translated text or to correct an error of execution” (p.330).

Briefly, the above discussion about writing processes has described the way writers compose texts, the different stages they engage in and their meaning making processes. Shaw and Weir (2007) contended that “[f]rom a cognitive perspective, a valid writing test would involve candidates engaging in all the processing components described above as appropriate to the level of proficiency being assessed” (p.42).

More specifically, the process of response validation was described by Alderson et al. (1995, p.176) as one in which the researcher seeks to “gather information on how individuals respond to test items. The processes they go through, the reasoning they

engage in when responding, are important indications of what the test is testing, at least for those individuals”. For instance, Wall et al. (1999) conducted a validation study for a new national examination in English in Sri Lanka and identified a number of limitations, like poor instructions, inadequate time control and unfamiliar test format. The students’ responses to the questionnaires indicated that the time was insufficient and the test was too difficult; therefore some items needed to be changed. McNamara (2000) was in agreement with this view and therefore stressed that feedback should be obtained from the test-takers through questionnaires, as test developers may find it difficult to spot the problems in the test. He pointed out that questionnaires should “include questions on perceptions of the level of difficulty of particular questions, the clarity of the rubrics, and general attitude to the materials and tasks” (2000, p.32).

#### **3.4.1.1.3 Scoring validity**

According to Shaw and Weir (2007, p.143), scoring validity “accounts for the extent to which test scores are based on appropriate criteria, exhibit consensual agreement in marking, are as free as possible from measurement error, stable over time, consistent in terms of content and engender confidence as reliable decision-making indicators”. In other words, scoring validity addresses the different elements of the testing situation that may cause a threat to the reliability of test scores. For instance, these elements may include: “[f]aulty criteria or scales, unsuitable raters or procedures, lack of training and standardisation, poor or variable conditions for rating, inadequate provision for post exam statistical adjustment, and unsystematic or ill-conceived procedures for grading and awarding” (ibid, p.143-4). Having identified these factors, it is important for examination boards to ensure as much as possible that none of these factors is present in any context

of assessment, otherwise the scores obtained from the test would have low reliability or low validity.

Weir (2005) identified a number of aspects of scoring validity for writing that include: criteria/rating scale, rater characteristics, rating processes, rating conditions, rater training, post-exam adjustment, and grading and awarding. It is important to consider these aspects in any testing situation as they have an impact on the reliability of test scores. This study will discuss the rating scale and rater training aspects as they are related to the question about scoring validity in this study.

#### **(a) Criteria/rating scale**

Shaw and Weir (2007) noted that there seems to be a consensus in the language testing literature that the validity of assessment in second language contexts is influenced primarily by the choice of appropriate rating scales and the consistency in applying them by trained examiners (cf. Hamp-Lyons, 1991b; Alderson, Clapham and Wall, 1995; Bachman and Palmer, 1996, McNamara, 1996 and Weigle, 2002). There are three main types of rating scales that are most referred to in the composition literature: primary trait scales, holistic scales, and analytic scales (Weigle, 2002).

#### **(1) Primary trait scoring**

This scoring method was developed by Richard Lloyd-Jones and Carl Klaus for the National Assessment of Educational Progress (NAEP) (Ruth & Murphy, 1988).

According to Davies et al. (1999, p.151), primary trait scoring “[i]nvolves the scoring of a piece of work (usually writing) in relation to one principal trait specific to that task”.

This definition shows clearly that primary trait scoring is context-specific. The rationale

for focusing on a specific context is to know how far candidates are able to elicit the kind of language that is appropriate to that context (e.g. argumentative or explanatory essays in academic contexts). Accordingly, different criteria need to be created and then applied for each task depending on the context of use. Odell (1981, p.124) noted that primary trait scoring “rests on the assumption that different tasks, even different expository tasks, may have to be judged by different criteria”. Thus, the challenge is in creating scoring guidelines that would reflect the qualities of writing which represent each writing task. More specifically, the scoring guide for each task includes

(a) the writing task; (b) a statement of the primary rhetorical trait (for example, persuasive essay, congratulatory letter) elicited by the task; (c) a hypothesis about the expected performance on the task; (d) a statement of the relationship between the task and the primary trait; (e) a rating scale which articulates levels of performance; (f) sample scripts at each level; and (g) explanations of why each script was scored as it was (Weigle, 2002, p.110).

In a nutshell, primary trait scoring states clearly in advance what is to be measured and how it will be measured. In doing so, it predetermines the primary trait for success in a task. Applebee (2000, p.96) pointed out that “[t]o insure that raters maintained this focus, scoring guidelines usually instructed raters to ignore errors in conventions of written language, and to focus on overall rhetorical effectiveness”. Nonetheless, since this rating scale is specific to a particular writing task and assessment, it is not possible to generalise to other types of writing task and it is time-consuming as well as expensive in terms of requiring detailed rating scales for each writing task (Shaw and Weir, 2007). Actually,

this is the reason why it is less widely used than holistic and analytic scoring in most assessment practices.

## **(2) Holistic scoring**

Davies et al. (1999, p.75) define holistic scoring as “[a] type of marking procedure which is common in communicative language testing whereby raters judge a stretch of discourse (spoken or written) impressionistically according to its overall properties rather than providing separate scores for particular features of the language produced (e.g. accuracy, lexical range)”. It is clear from this definition that giving a holistic score to a piece of writing requires raters to read the whole essay in order to have an overall impression of its ‘global’ quality and neglect the specific linguistic features of the written text. During the scoring process, raters refer to a scoring rubric that consists of a number of criteria exemplifying the kind of language expected at each level of proficiency. For the sake of producing consistently reliable scores, “efforts should be made to socialize a rater to the ways other raters assign grades and their rationales for doing so” (Grabe and Kaplan, 1996, p.404).

The major advantages of such methods include: (1) rapid scoring of texts, (2) more economic to mark than any other method and (3) the rater focuses on the strengths of a piece of writing rather than its shortfalls (Shaw and Weir, 2007). This could be why it is the most common instrument used for large scale writing assessment (Grabe and Kaplan, 1996). However, holistic scoring has received a number of criticisms in the literature. Hout (1990) summarised these drawbacks as being:

(1) that holistic ratings correlate with appearance and length (Markham; Sloan and McGinnis); (2) that the product orientation of holistic ratings is unsuitable for informed decisions about composition instruction or student writing (Faigley et al.; Gere; Odell and Cooper); (3) that holistic ratings cannot be used beyond the population which generated them, so holistic scoring is useless as an overall indicator of writing quality; and (4) that holistic training procedures alter the process of scoring and reading and distort the raters' ability to make sound choices concerning writing ability (Charney; Gere; Huot). (p.201-2)

In addition, the interpretation of scores is not always easy especially when raters tend to give the same score but use different criteria for scoring (Weigle, 2002). That is to say, different raters may apply different criteria when scoring the same piece of writing and arriving at the same score. In fact, this variation of scoring threatens reliability and hence it is suggested that more than one rater should be involved and their judgments pooled (Davies et al., 1999).

### **(3) Analytic scoring**

Analytic scoring was first developed by Diederich through his work on direct assessment of writing (Hout, 1991). Britton et al. (1966) traced earlier attempts (Rice 1903; Hillegas, 1912; Willing, 1918; 1926; Val Wagenen, 1920) to establish analytic scales in order to enhance the consistency of scoring compositions. Analytic scoring refers to those assessment practices “where a separate score is awarded for each of a number of features of a task, as opposed to one global score” (Davies et al., 1999, p.7). More specifically, the features of the text, in a writing test, that assessors consider in scoring may include content, coherence and cohesion, organisation, and accuracy. As a result, analytic scoring

provides a richer account of the candidate's performance in different criteria of writing. Unlike holistic scoring where a general impression is given to the whole text, analytic scoring as the name suggests can assist raters in addressing each aspect of the written text and thus a balance is created among them in terms of measuring against the same criteria. Weigle (2002) listed a number of advantages of analytic scoring (see Table 6 for a comparison between holistic and analytic scoring):

- 1- Analytic scoring is more useful in rater training, as inexperienced raters can more easily understand and apply the criteria in separate scales than in holistic scales.
- 2- Analytic scoring is particularly useful for second language learners, who are more likely to show a marked or uneven profile across different aspects of writing
- 3- Analytic scoring can be more reliable than holistic scoring. (p.120)

<b>Quality</b>	<b>Holistic Scale</b>	<b>Analytic Scale</b>
Reliability	lower than analytic but still acceptable	higher than holistic
Construct Validity	holistic scale assumes that all relevant aspects of writing ability develop at the same rate and can thus be captured in a single score; holistic scores correlate with superficial aspects such as length and handwriting	analytic scales more appropriate for L2 writers as different aspects of writing ability develop at different rates
Practicality	relatively fast and easy	time-consuming; expensive
Impact	single score may mask an uneven writing profile and may be misleading for placement	more scales provide useful diagnostic information for placement and/or instruction; more useful for rater training
Authenticity	White (1995) argues that reading holistically is a more natural process than reading analytically	raters may read holistically and adjust analytic scores to match holistic impression
Interactiveness*	n/a	n/a

Table 6: A comparison of holistic and analytic scales on six qualities of test usefulness (Adapted from Weigle, 2002, p.121)

Above all, analytic scoring is designed to assess the predefined qualities of the written text which are crucial for the context of assessment. While analytic scoring has proved to be more reliable than holistic scoring it has been criticised in the writing assessment literature for being “uneconomical, unreliable, pedagogically uncertain or destructive, and theoretically bankrupt” (White, 1985, p.124). It is uneconomical because the scoring process is time-consuming as it requires the raters to go through each criterion in order to score the written text (Weigle, 2002). It is destructive and theoretically bankrupt because it sometimes focuses on “relatively trivial features of text (grammar, spelling, handwriting) and which did indeed reduce writing to an activity apparently composed of countable units strung together” (Hamp-Lyons, 1991b, p.247) which may as a result “distort and misrepresent the writing process” (Marsh and Ireland, 1987, p.8). Actually, unless the criteria in the analytic scales are clearly defined they are otherwise of little use to the stakeholders (Braddock et al., 1963).

Furthermore, it has been reported that analytic scoring (or multiple trait scoring as some researchers would like to call it, though other researchers do not treat them as synonyms for each other, e.g. Hamp-Lyons, 1991b) gives rise to a halo effect because “when raters are asked to make multiple judgements they really make one and this affects all other judgements” (Fulcher, 2010). Similarly, during the scoring process raters’ attention may be diverted from the overall effect of aspects of performance due to focusing on specific ones but “this problem may be at least partially overcome by requiring raters to give an overall impression score in addition to the analytic scores” (Davies et al., 1999, p.7).

To sum up, it has become clear from the above discussion that each scoring method has both advantages and disadvantages and thus the use of any method should go through a careful analysis of its merits in order to judge its appropriateness for the context of use. This reflects the complexity of writing and of its assessment as a multifaceted practice. Analytic scoring seems to be the most suitable method that attends to some extent to the different facets of the written text and thus, if designed properly, provides more reliable and valid diagnostic accounts of an individual's performance in writing.

For the purpose of this study, it is worth pointing out that the IELTS AWM uses an analytic scoring approach for the assessment of the written performance of candidates. To overcome the above problems of analytic scoring, the rating scales for IELTS AWM have been validated in a number of phases where they also proved to have high reliability (see Shaw, 2002a; 2002b; 2003a; 2003b; 2004; 2005). In addition to the above advantages, analytic scoring is “more suitable for second-language writers as different features of writing develop at different rates” (Shaw and Weir, 2007, p.151) and this may be the rationale for IELTS preferring analytic scoring to any other method.

#### **-Empirical studies on the application of rating scales**

Scoring written texts in order to arrive at one score is a complex activity. One way to better understand this complex practice and get to know the extent to which rating scales are used when making decisions during the scoring process is to ask raters to reflect on their cognitive activities (thinking processes) as they score a piece of writing in what is known as verbal or think-aloud protocols (Green, 1998). Sakyi (2000) investigated the

thinking processes and criteria used by raters in evaluating written texts holistically. His analysis revealed the different reading styles adopted by raters as they read the texts: “(1) focus on errors in the text, (2) focus on essay topic and presentation of ideas, (3) focus on the rater’s personal reaction to text, and (4) focus on the scoring guide” (p.129). This indicates that not all raters focus on the scoring guide as there were other aspects which diverted their attention from it. For those who deliberately followed the scoring guide, they were obliged to depend on one or two specific features to differentiate between various levels of ability. He suggested a tentative model of the different factors that affect the holistic scoring process. These include: content factors, language factors, raters’ personal biases/expectations and personal monitoring factors.

More specifically, in a review of empirical research on factors that influence holistic rating, Charney (1984) identified a number of superficial and mechanical characteristics of the written text which are irrelevant to the actual performance of candidates in writing. Raters did not adhere to the scoring criteria as they were affected by the following factors: physical appearance (poor handwriting tends to be associated with low holistic scores), word choice, length of essay and spelling errors. She commented that “[i]t is disconcerting to find holistic scores, which are supposed to be a qualitative measure, so directly predictable by such mundane quantitative measures as the length of the sample, the number of errors and the number of unusual vocabulary items” (p.75). In other words, holistic scoring according to this view (which is also dominant in the language testing literature) should be concerned with the overall quality of the text rather than quantity i.e. counting the specific features of the text. Similarly, Brown (1991) examined the

performance of raters on a 6-point holistic rating scale. His feature analysis of raters' behaviour indicated that there may be variations in perspective when awarding final scores.

However, the study by Lumley (2002), in which he investigated the processes by which 4 trained raters made scoring decisions when using an analytic scale to score 48 written texts by ESL learners, found that raters used similar rating processes in three stages. Nonetheless, they found it was difficult to reconcile their impression of the text, the specific features of the text, and the wordings of the rating scale. He suggested that raters can be successful in yielding consistent scores if they receive adequate training and assistance in the form of guidelines to deal with problems arising during the scoring process (ibid). Weigle (1994) is perhaps in accordance with this view as she explored the effect of training on 4 inexperienced raters of ESL compositions before and after the training. Her study demonstrated that training made the scoring criteria clear for raters, changed the way they perceived students' writing and provided a reference group of other raters to compare their performance against.

The above few examples from the writing assessment literature have briefly shown the misuse of rating scales, lack of experience, and the variation in judging written performance. No matter what scoring method is used, it seems that the more the raters diverge from the scoring guide the less reliable and less valid their assessment. That is to say, low inter-rater reliability results from giving different scores to the same piece of writing and low validity comes from using an instrument which does not fully represent

the construct. The different interpretation practices of candidates' performance on the part of the raters and the misuse of scoring criteria together affect the reliability and the validity of test scores. This led McNamara (1995) to wonder doubtfully "whose performance are we assessing?" and he called it the 'Pandora's Box' of language testing (Quoted by Alderson and Banerjee, 2002, p.81). In doing so, he required language testers to take into account the social dimension of language testing and its relative factors which had been ignored in the past. Specifically, such a comprehensive view of assessing performance would include the interactions between the task, the test taker and the rater (see Figure 2). Although the figure below is about the interactions in the assessment of speaking, this can also be applicable to the assessment of writing in which interactions take place indirectly.

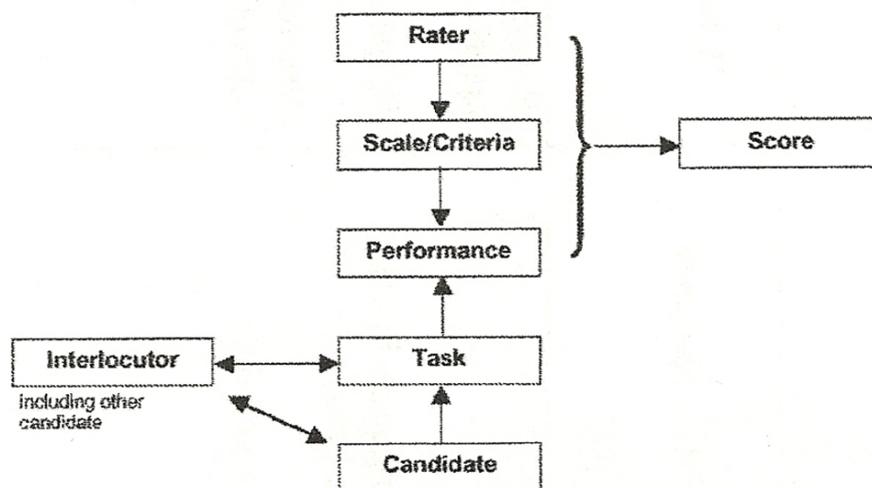


Figure 2: Interactions in performance assessment of speaking skills (Adapted from McNamara, 1995)

### (b) Rater training

The consistency in applying scoring criteria helps us to obtain high scoring validity through intra-rater reliability (self-consistency) and inter-rater reliability (agreement

between raters) (Weigle, 2002). Nonetheless, she found that “rater training is more successful in helping raters give more predictable scores (i.e., intra-rater reliability) than in getting them to give identical scores (i.e. inter-rater reliability)” (Weigle, 1998, p.263). That is to say, inter-rater reliability is not easy to achieve in the assessment of written performance. In doing so, she seems in agreement with Barritt et al. (1986) when they argued that:

To require evaluators to pay attention only to pre-established criteria is to imply that this is the only way for fair readers to act. Such a stance makes reading very difficult for anyone who wants to take seriously her own response as an expert, particularly as an expert teacher, while at the same time trying to reach consensus with other readers...When reading is understood as re-creation of a writer's intended meanings, then it is clear that there will often be more than one reading of a text, for there is no good reason why two different readings might not both be accurate ones (p.324).

For this reason, rating scales which have been designed for the sake of ensuring reliability and validity among raters as referring to one common standard may not be sufficient alone. Shaw and Weir (2007) suggested that in order to convey a standard we need not only depend on written assessment criteria but also on examiner training and standardisation. Although training examiners to be consistent in terms of harshness and to eradicate any other differences among them may be “unachievable and possibly undesirable” (McNamara, 1996, p.232) because each rater has his or her own unique way of perceiving the task and the scoring criteria, it is still possible to “bring examiners’ differences in severity to a tolerably acceptable level” (Lunz, Wright and Linacre, 1990; cited in Shaw and Weir, 2007, p.181). According to Bachman and Palmer (1996) this can be done through a training programme which includes the following six steps:

- 1- Read and discuss scales together.
- 2- Review language samples which have been previously rated by expert raters and discuss the ratings given.
- 3- Practise rating a different set of language samples. Then compare the ratings with those of experienced raters. Discuss the ratings and how the criteria were applied.
- 4- Rate additional language samples and discuss.
- 5- Each trainee rates the same set of samples. Check for the amount of time taken to rate and for consistency.
- 6- Select raters who are able to provide reliable and efficient ratings. (p.222).

To sum up, building on these three types of validity, this study will examine the tasks of the Academic Writing Module (AWM) in the IELTS test from the perspectives of the stakeholders in terms of clarity, length, style, level of difficulty, balance of themes, suitability to the learners, authenticity and assessment. It will also investigate how students respond to these tasks and the extent to which this affects their performance. Furthermore, it will investigate the extent to which the rating scales are used consistently by examiners.

### **3.5 Empirical studies on IELTS**

#### **3.5.1 Predictive studies on IELTS as a preparation to academic study**

Studies on the predictive validity of IELTS have found mixed results. Ferguson and White (1998) conducted a small scale study over an academic year at the University of Edinburgh in the United Kingdom. The researchers found that there was a weak positive relationship between language proficiency and academic outcome. Their scores on the initial test correlated 0.39 with academic outcome and 0.46 on re-testing. The researchers found that there were a number of factors that contributed to academic failure. These included: academic ability, language proficiency, personal circumstances and traits. The study pointed out that these factors were the result of intensive study in a one year Masters course, academic writing in exam conditions, adjustment to a different culture and problems concerning mixed groups of home and overseas students.

Hill et al. (1999) carried out a study at the University of Melbourne and found that there was a moderately strong relationship ( $r = 0.540$ ) between Grade Point Average (GPA) and IELTS scores while there was a weak correlation between achievement and TOEFL scores. Their study investigated the usefulness of IELTS and TOEFL as predictors of readiness for the Australian academic context. Likewise, Feast (2002) examined the relationship between IELTS scores and academic performance as measured by GPAs at one South Australian university. In this study, she found that there was a positive but weak relationship between IELTS scores and academic success.

Most importantly, Ingram and Bayliss (2007) investigated the relationship between IELTS scores and students' 'behaviour' in a university context. The study found that

students produced the kind of language expected by the IELTS score in contexts related to their academic courses. Yet, the study indicated that the students' performance in academic contexts which were beyond the scope of the IELTS test had no relationship with their IELTS scores. In addition, findings from the study revealed that a small number of students who had satisfied entry requirements were not able in terms of their language to meet the demands of the academic tasks in the programme. As a result, this has led the researchers to question the adequacy of entry level requirements in such courses.

In contrast, Cotton and Conrow (1998) at the University of Tasmania in Australia found no positive correlations between IELTS scores of the student group under investigation and three measures of academic achievement: GPAs, academic staff ratings of student performance and students' self-ratings of performance. In addition, they found no positive correlations between IELTS scores and language difficulties related to coursework. Yet, there was a weak positive correlation in the reading and writing scores at 0.36 and 0.34 with staff ratings of academic achievement, and 0.46 and 0.39 with students' self estimates of academic performance in the second semester. The study indicated that language difficulties were not the only variables affecting academic performance. Other variables were, namely, the amount of English language tuition received, motivation, cultural adjustment and welfare difficulties experienced by international students.

Similarly, the results from Kerstjens and Nery's (2000) study in Australia indicated a weak positive correlation between reading (0.262) and writing (0.204) tests and GPA. While the total sample and the Higher Education group had a small to medium predictive effect on academic performance from the IELTS scores, there was no significant

prediction of academic performance for the Technical and Further Education (TAFE) group. The study pointed out that there are many factors that can affect the academic outcomes of international students. These include: language and socio-cultural and psychological factors like learning and educational styles, social and cultural adjustments, motivation and maturity, financial and family pressures.

In support of this view, together with other researchers such as Gibson and Rusek (1992) and Bellingham (1993), Dooley (1999) at Curtin University of Technology in Australia found that language plays a less significant role at the higher levels of proficiency. Thus, students who do not meet the English language requirements (6 on IELTS) can succeed academically because language is only one factor among many that influence academic performance. Though the findings of this study did not “offer conclusive evidence about the validity of IELTS”, the reading module of the test showed a strong correlation with academic success (Dooley, 1999, p.114).

### **3.5.2 Studies on tasks design for IELTS academic writing**

Moore and Morton (2007), in Australia, investigated the authenticity of Task 2 in the IELTS academic writing module and examined its relevance to university writing. They found that the genre was somehow similar in both academic and general domains. Yet there were some differences, such as:

1. The use of prior knowledge as the basis for writing in the IELTS tasks, compared with the prescription of a variety of research processes in the university assignments.

2. A restricted range of rhetorical functions in the IELTS items (with a focus on hortation), compared with a diversity of functions in the university tasks.
3. An emphasis on ‘real world’ entities (situations, actions, practices) as the objects of enquiry of IELTS items compared with a greater focus on abstract entities (theories, ideas, methods) in the university tasks. (Moore and Morton, 2007, p.197).

The study also found that the IELTS tasks are opinionated in terms of style which makes them more related to specific public non-academic genres. Consequently, the researchers suggested that these types of tasks are not a suitable model for university writing.

Banerjee (2000) seems to be in total agreement with Moore and Morton’s (2007) argument, as she contended that IELTS academic writing modules are relatively short compared to university assignments. Hence, they fall short of providing stakeholders with the necessary information about the ability of the candidates to produce coherent, appropriate and accurate texts which are similar in terms of length to those produced at university (See Table 7).

<b>Task</b>	<b>IELTS</b>	<b>Research Degrees</b>
Writing Task 1: “Describing information” (graph/ table/ chart diagram)”	Focuses on - Organizing, presenting and comparing data; or - Describing the stages of a process or procedure; or - Explaining how something works No speculation on the reasons for data, procedures or the way something work	Focuses on - Organizing, presenting and comparing data; or - Describing the stages of a process or procedure - Explaining how something works and; - Many other tasks All data incorporated into a “critical approach” Reasons suggested for data and procedures critiqued
Writing Task 2: “Essay”	Presenting a written argument or case on a given topic Information only from the prior knowledge No referencing, conclusions based purely on anecdotal evidence Focus on moral/ social desirability of practices Focus on concrete objects of enquiry	Many types of writing including “essays”, case study reports, research reports etcetera Information from primary or secondary research sources Appropriate citation vital Focus on scientific validity or summarising views related to practices Focus on both concrete and abstract objects of enquiry

Table 7: Tasks and Skills in IELTS versus those in Research Degrees (Adapted from Picard, 2007, p.445).

O’Loughlin and Wigglesworth (2007) in Australia investigated task 1 difficulty in the academic writing module in the IELTS test. Task difficulty was examined in terms of amount and presentation of information to candidates. The researchers found that tasks with less information were more complex than those with more information. They indicated that the more information candidates have in tasks the more accurate the responses will be. They suggested the use of simpler tasks for achieving higher performances from candidates.

However, O'Loughlin and Wigglesworth's (2007) study was limited to three types of graphs and two topics. Yu et al.'s (2011) study was more comprehensive in terms of including all types of graphs and 8 academic writing tasks with a variety of topics. In this study, they investigated the cognitive process of candidates taking IELTS AWM task one. The researchers gave different graphic prompts pre and post-training on how to answer the task.

They found that the graphic conventions of graphs affected the processes of comprehending and re-producing graphic information. Moreover, the candidates pointed out that graph familiarity had some potential psychological impact on their task performance. Furthermore, the researchers noticed that the participants used different vocabulary choices for different graphic prompts as a result of the effects of the writing abilities on their cognitive processes. Also, training had a strong impact on the performance of the candidates. The researchers suggested that:

- 1- When designing academic writing tasks one in IELTS, we need to consider what accompanying instructions should be provided with the graphs and whether the summary-like introductory sentence should serve as the entry point for leveling the playing field for test takers with varying background knowledge, graphic skills and reading abilities.
- 2- As academic writing tasks one are essentially reading/writing integrated tasks, it is essential to consider not only the graphic skills of the test takers but also their reading abilities.
- 3- It is also important to consider and compare the potential differential effects of different graphic prompts on the cognitive processes of test taking as well as on

test performance, because different graphs have different conventions in presentation and interpretation.

- 4- We also need to consider the information density, or data points of the graphs, as information density clearly has impact on how and what test takers can extract from the graphs.
- 5- The quality of a graph is not just innate in its display features. It is in fact determined in interaction between these features and the characteristics of graph readers (e.g., their familiarity with and experience in using graphs) (Yu et al., 2011, p.35).

Task 2 was explored by Mayor et al. (2007) with respect to the effect of topic or wording on candidates' writing. They indicated that candidates tended to use specific linguistic forms due to certain cues found in the wording of the prompt in the task. This effect of the prompt on the response was clearly observed by researchers through the overuse of interpersonal reference and limited use of impersonal language. In addition, they found that "the nature of topics could have led to differential gaps in candidates' knowledge of the relevant vocabulary" (ibid, p.298).

Similarly in Australia, Mickan et al. (2000) carried out a response validity investigation on the test prompts of the IELTS writing test with specific focus on their 'readability' to candidates in terms of discourse and pragmatic features. They found that the purpose of the tasks as well as the lexicon-grammar of the tasks have an influence on candidates' understanding of the test prompts. In addition, findings of another study revealed that different interpretations of the test prompts have led, as a result, to different responses (Mickan, 2003). Furthermore, Mickan and Slater (2003), in a similar investigation of

academic writing tasks, found that anxiety and time constraints made the process of text production more complex. Successful candidates were those who were quick in interpreting the prompts in order to understand the topic and purpose of the task and thus could finish the task within the allotted time. The researchers asserted that candidates should not be blamed for not meeting the rhetorical expectations of Task 2 as it was not sufficiently clearly described in the IELTS Handbook.

### **3.5.3 Studies on assessment of IELTS academic writing tasks**

In examining the problem of inconsistency in rating IELTS exams, Mickan (2003) called for the need to have valid criteria for the assessment of academic writing tasks.

According to him, determination of written performance is a complex semiotic process that is carried out by candidates who interpret the prompts and write texts accordingly, and on the other hand by raters who interpret rating criteria, analyse scripts and assign the corresponding score to the students' level of proficiency (Mickan, 2003). His analysis of texts revealed that identifying lexico-grammatical features that distinguish levels of performance was hardly achievable. Therefore, the study suggested that scoring of written texts should be holistic rather than analytic.

In contrast, Rignall and Furneaux (2002, cited in Green, 2007a) investigated IELTS trainee raters prior to and following a rating training programme and found more consistency among raters when using an analytical scale than when applying global rating. Likewise, Green (2007a) reported high inter-rater consistency among IELTS AWM raters when using analytic scoring while global scoring showed low inter-rater consistency. This argument may tell us why the IELTS test adopts an analytic scoring for the AWM. However, though the IELTS claims the use of analytic scoring (IELTS

Handbook, 2007), Mayor et al. (2003), in a study about the design and content of the IELTS AWM test in terms of the linguistic features of candidates' output, argued that raters tended to apply holistic rather than analytic scoring when responding to scripts. They also found that assessment of AWM in the test is different from that of a tertiary level as candidates are dependent on their own resources, a situation which does not echo academic writing norms at university.

Fulcher (2009) seems to support Mayor et al.'s (2003) argument about IELTS scoring as he took it further with relation to Thorndike's (1920) concept of the Halo Effect in scoring. Fulcher contended that raters in tests like IELTS that uses multiple trait scoring (traits in IELTS AWM include: task achievement or response, coherence and cohesion, lexical resource or grammatical range, and accuracy) do not make multiple judgements but one as they are affected by the first grade which they assign to the first trait. He concluded that profiles tend to be 'flat', defeating the aim of providing informative, rich information, on learner performance (Fulcher, 2009). Rignall and Furneaux (2007) appear to be in line with this view when they explored the effect of standardisation-training on rater judgements for IELTS AWM. They found that raters tended to neglect certain criteria, such as coherence and cohesion, when scoring candidates' scripts, even after completing a training programme.

#### **3.5.4 Impact and washback studies on IELTS and IELTS preparation courses**

The terms 'impact' and 'washback' have sometimes been used synonymously, but Bachman and Palmer (1996) believed that impact is wider and takes place at the macro or social and institutional level, while washback occurs at the micro or individual level. In addition, the terms washback or backwash have been used interchangeably in the

literature to refer to the same phenomenon (Bachman, 1990, Alderson and Wall, 1993, Davies et al., 1999, Hughes 2003). Alderson and Wall (1993), who agree on using these terms interchangeably but prefer washback, traced their use and found that backwash was used in general education articles while washback was used in British applied linguistics. Washback can be positive or negative and it is defined as “the extent to which the test influences language teachers and learners to do things they would not necessarily otherwise do because of the test” (Hughes, 1994, p.117). This phenomenon of washback on IELTS, either with positive or negative impacts, is under-researched and there are few empirical studies in the literature.

The earliest and largest study which investigated this phenomenon in depth on IELTS was a longitudinal study called “*IELTS Impact Study*” which was launched in 1995 (Hawkey, 2004). Major findings from this impact study indicated that 90% of teachers agreed that the IELTS test influenced the content of their lessons while 63% of them agreed that it influenced their methodology. In addition, the study found that IELTS had a significant impact on preparation courses. Furthermore, one-third of the students felt “very worried” before taking the test. This was supported by 53% of teachers believing that the test caused some stress. Yet, 94% of teachers claimed that the test provided a positive motivation for candidates.

However, this study addressed the macro level, i.e. impact, which has to do with society or social institutions. Anthony Green’s study has addressed the micro level, i.e. washback, where the individual is the focus of attention. His study on IELTS washback was a PhD study conducted between 2001-2004, in the United Kingdom, on “*Test Impact and EAP: a comparative study in backwash between IELTS preparation and university*

*pre-sessional courses*". It was published in *Studies of Language Testing* (2007a) vol. 25. In this study, he investigated the influence of the IELTS Academic Writing Module on preparation for academic study and the efficiency of IELTS test preparation and other forms of English for Academic Purposes (EAP) intended for university study. It was found through inferential statistics that there was an improvement of scores through instruction in academic writing and there was no significant difference in terms of score gains between those students who were enrolled in EAP pre-sessional courses and those in IELTS preparation courses. In other words, there was "no clear advantage of focused test preparation" (Green, 2007b, p.75). Moreover, students who planned to take the test again as well as those who had low initial writing test scores were the only two groups of learners who showed improvements in terms of score gains. Such improvement in test scores was seen to have greater relevance to washback to the learner through motivation to succeed rather than washback to the programme (Green, 2007b).

Green (2007a) seems to argue against the results produced by Brown (1998) who claimed that students in the IELTS test preparation courses tended to have higher scores than those in the EAP courses. In the IELTS preparation course, students' average gain was 0.94 of a band on the Academic Writing Module (AWM). In contrast, students in the EAP course whose programme was not IELTS oriented experienced a decline in their scores by 0.6 within the same period. The study found that the IELTS preparation course was more concerned with preparing students for the test tasks rather than equipping them with research skills and writing techniques required by university study. Brown suggested the need for a larger sample as these findings from his small scale enquiry cannot be generalised due to sample size. He also highlighted the fact that students in the IELTS

preparation course may have stronger motivation as they were going to take the test immediately after the course whereas the EAP students had plenty of time before the commencement of their university programmes.

Unlike the small scale study conducted by Brown (1998), Deakin (1996) explored washback from the IELTS test on EAP course providers in Australia. Deakin found that more than 50% of the surveyed centres offered IELTS preparation within EAP. In addition, 29% had test-focused preparation courses. Most participants responded positively and considered the IELTS as a leading path to EAP preparation. Yet 37% of the respondents to the initial survey highlighted the negative washback effect of IELTS on EAP teaching and university preparation. Results from the qualitative follow-up survey and from case studies supported this finding and indicated that IELTS preparation programmes failed to cover EAP issues as they were teaching to the test. More specifically, the case study results shed light on some significant skills required by university study which the IELTS test ignored. These include:

- understanding subject-specific concepts
- specialised language and vocabulary
- interpreting assessment tasks
- time management and organisational skills
- research skills
- cognitive skills/critical skills
- understanding academic requirements

- cultural understanding: interacting with Australian students
- citing references in writing/plagiarism
- seminar presentations
- organising and writing longer papers
- participating in tutorials
- communicating with lecturers/supervisors (Deakin, 1996).

Hayes and Read (2004) in New Zealand supported the finding that there exists a washback from IELTS preparation courses, as teachers teach to the test and thus fail to address EAP issues. Interest in the study started when the researchers felt that students in the preparation programmes did not have sufficient linguistic ability to deal with academic language. Unfortunately, even those who satisfied entry requirements to university programmes were characterised by their teachers as having poor command of the academic language needed at university level. The study concluded that there was an obvious washback effect in the IELTS preparation courses as the teacher and students were more concerned with practising the tasks of the test rather than developing the academic language proficiency needed for university study. In other words, the IELTS preparation courses prepared students for the test and not for the academic language that they were going to encounter at university.

For this reason, Everett and Coleman (2003) thought that the issue of washback in IELTS might also exist in preparation materials. They examined six IELTS preparation materials in three Australian language centres with a specific focus on the appropriateness of

content, organisation, learning approaches and presentation of the listening and reading components. The researchers indicated that some teachers had students who found that there was a missing link between IELTS preparation and academic study. They proposed that inclusion of more texts and tasks in IELTS preparation materials would contribute to the social and academic acculturation of students.

In contrast to the studies above, Rea-Dickins et al.'s (2007) study looked at the post-IELTS impact. The researchers examined the possible affective and academic impacts of the IELTS performance of a postgraduate group in the United Kingdom. Interestingly, they explored the language proficiency of learners with respect to learning experience, academic profession and identity. This study has a number of findings which point to:

- (i) the affective dimensions of the struggles of postgraduate students and the ways in which these derive from the test itself;
- (ii) the linking of this struggle with how they work through the four language skills;
- (iii) an overwhelming lack of awareness of admissions staff about IELTS;
- and (iv) the assumptions about the test by tutors and how these might impact on student performance (Rea-Dickins, et al., 2007, p.59).

The above studies have tackled the phenomenon of impact or washback with focus on either stakeholders (Hawkey, 2004 and Rea-Dickins et al., 2007) or preparation courses and materials (Green, 2007a; Brown, 1998; Deakin, 1996; Read and Hayes, 2004 and Everett and Coleman, 2003). Generally, the present study supports the above literature. Previous studies have been fragmented in terms of addressing the issue of validity of the IELTS AWM (i.e. focusing on one type of validity). Therefore, this study argues that the validity of the IELTS AWM may be affected by one or more of these three factors, namely the test, the candidates and the markers. With respect to the content validity of the

test, this study contends that the tasks in the writing test may not reflect the kind of tasks candidates are likely to encounter at university. Moreover, this study asserts that response validity, on the part of students who may not have understood the rubric of the tasks, seems to be another factor affecting the students' performance. Furthermore, this study argues that scoring validity may affect the students' scores in terms of markers who may have failed to assign the students to their corresponding level of proficiency. In considering all of these types of validity which together constitute construct validity, this study may inform us why most Saudi candidates fail to cope with the demands of university study despite the fact that they have achieved the minimum entry requirements in IELTS.

### **3.6 Research Questions**

This study has set out to investigate the following question that has three specific sub-questions under it, addressing each type of validity.

How do stakeholders perceive the construct validity of the academic writing module in the IELTS test?

#### **Content Validity:**

1- In what ways do academic writing test tasks correspond to the types of tasks students do at university?

#### **Response Validity:**

2- a- What sense do students make of the tasks in the writing test?

b- To what extent does this affect their performance?

### **Scoring Validity:**

3- To what extent are the rating scales used consistently by markers?

### **3.7 Summary and Conclusion**

This chapter has reviewed the most important issues in the literature related to the focus of investigation. It highlighted the different theories of writing composition (process and product) and assessment with special reference to direct assessment of writing in second language contexts. In doing so, it presented the socio-cognitive framework underpinning the study which regards the composition and the assessment of writing as being both cognitive and social. Because of the inseparability and the complex relationship between context and cognition in interactions that bound the writer, the process (task), the product (text), the reader and the context, this study made use of five approaches in order to arrive at a comprehensive description of the above elements that may affect candidates' performance in the IELTS Academic Writing Module (AWM). These approaches included: writing as a cognitive practice, a social construction practice, a social interaction activity, and as a socio-cultural practice.

More specifically, these approaches may have some links to IELTS AWM in the following ways. On one hand, because of the time constraints candidates sitting for the IELTS AWM test may consider aspects like time management as top priority over a post-hoc revision or other processes. For this reason, timed tests of writing are scantily suited to the assessment of the process of writing as they assess the written product and thus give little evidence on how the text was composed (Hamp-Lyons and Kroll, 1997). Moreover, low level tasks (e.g. IELTS Task 1) that had narrative or instructional texts

required knowledge telling skills, while high level tasks (e.g. IELTS Task 2) that involved argumentative texts, demanded knowledge transforming skills (Shaw and Weir, 2007). On the other hand, candidates' writing in the IELTS AWM test can be considered as a result of complex social interactions shaped by the context. Their texts were a response to the tasks' prompts which were designed by the test developer in order for the candidates to meet the expectations of the markers and the specifications of the test itself. It can be seen here that the interest was in the production of texts for the sake of assessment rather than for communicating information. Yet, IELTS AWM test tasks can be thought of as being interactive as they can engage test takers in almost all the metacognitive processes characterising interactive tasks. For example, they can allow students to 'self-mediate' their interactions with the test tasks in their attempts to retrieve internalised external social interactions. Actually, an individual's score in the test may represent not only his/her cognitive or metacognitive processes but also internalised social activities mediated through language.

This chapter has also shed some light on the developments in the field of language testing and assessment which have responded to shifts in paradigms in language teaching and learning theories (e.g. subjective testing, scientific (objective) testing based on structuralism and behaviourism which were then replaced by audiolingualism and recently by communicative testing) since the middle of the last century. These theoretical developments have given rise to a number of test types but this study discussed the most frequently used tests in the literature and which were suggested by Alderson et al. (1995). For instance, these tests included: placement tests, progress tests achievement tests, proficiency tests and diagnostic tests. Then, special reference was made to the IELTS test

which is a proficiency test. As IELTS is the focus of the study a review of its development, content, format, and marking was felt to be necessary for an understanding of its construct and specifications.

This review has discussed the concept of validity in language testing field because it is the phenomenon under investigation in this study. It mentioned different types of validity with special focus on three types of validity examined in the study, i.e. content validity, response validity and scoring validity. It was made clear that this study considered these three types of validity as constituting a unitary concept, i.e. construct validity (Messick, 1989, 1996a, 1996b). They were seen as dependent on each other and treated separately throughout this study for descriptive purposes only.

At the end of this chapter, empirical studies on IELTS which were at the heart of the aforementioned three types of validity were reviewed because they were the focus of enquiry. Namely, they were: predictive studies on IELTS as a preparation to academic study, studies on tasks design for IELTS academic writing, studies on assessment of IELTS academic writing tasks and impact and washback studies on IELTS and IELTS preparation courses. After this extensive review of pertinent literature on empirical studies about the validity of the IELTS test and its AWM, this chapter has concluded with the research questions that identify the phenomenon under investigation.

This study is different from those reviewed in the literature in the following ways:

- 1- It is investigating both of the writing tasks from a socio-cognitive perspective which recognises that writing is both a cognitive process and an act of socially situated meaning making.

- 2- It is approaching the problem from a construct validity approach (Messick, 1989) which unifies the three types of validity under investigation, i.e. content validity, response validity and scoring validity.
- 3- It is triangulating the data in a qualitative enquiry comprised of three different data collection methods, i.e. open-ended questionnaires, interviews and observations.
- 4- It is adopting a multi-disciplinary theoretical perspective in relation to writing drawing on cognitive psychology and socio-cultural theories of writing.
- 5- There are three stakeholders represented in the study, i.e. candidates, markers, and university lecturers.
- 6- It is approaching the problem through a case study that followed the same participants on two occasions.

Put together, these elements, to the best of my knowledge, cannot be found in one study investigating the construct validity of the Academic Writing Module (AWM) in the IELTS test. This is where this study locates itself in the literature and this is what it is trying to achieve in the following chapters.

# CHAPTER 4

## **Chapter IV**

### **Methodology**

#### **4.1 Introduction**

This chapter describes the methodology adopted in this study. It starts by giving a brief description of the philosophical assumptions underlying the current study by outlining the ontology and epistemology of the informing paradigm. It then explains the rationale for choosing the qualitative approach. This is followed by the researcher's role, the method of sampling, and the context and the participants of the study. The main focus is on describing the data collection procedures which include the design, the piloting and the administration of each method used in the study. It then sheds some light on data analysis, and ethical issues. The chapter concludes with a summary of the research process followed in this study.

#### **4.2 Ontology**

Philosophical assumptions are essential in guiding all research, whether quantitative or qualitative. Therefore, this section will discuss the underlying assumptions of this enquiry. The interpretative paradigm of this study adopts a relativist ontology which assumes that there are multiple realities (Denzin and Lincoln, 2005). Crotty (1998, p.64) stressed that "we should accept that social constructivism is relativist". Lincoln and Guba (1989) who also assign relativist ontology to the interpretative paradigm, refer to the constructivist paradigm to mean interpretative (Ernest, 1994). According to him,

relativism is “the view that knowledge is dependent on the knower and his/her context, possibly on a community of knowers and their social context, but cannot be objective or absolute, transcending humanity (Ernest, 1994, p.36). In this study, the obtained knowledge is dependent to a large extent on both the researcher and the participants, and their contexts as well. Such knowledge which emerges from the contexts is to be considered in relation to each resource. Most importantly, the judgements that the researcher makes about the participants are mainly derived from the participants’ points of view about reality which are collected in a natural setting. These viewpoints may be further verified, for the sake of validation and sorting out differences, by those viewpoints that the researcher has come to notice in the context. For example, data from observations or questionnaires can be understood by comparing them with data from interviews.

In this sense, the researcher is engaged in an endless process of interpretation while attempting to figure out the complexity of the context through interconnected viewpoints of participants. Hodkinson and Smith (2008, p.154) asserted that “[r]elativists, as constructors of knowledge, must hold that while certainly the circle of our interpretive discourse may expand and deepen and become more interesting or even more useful, there is no way out”. Nevertheless, in this process of interaction between the researcher and the participants, findings are created not discovered by the addition of meaning to the phenomenon (Pring, 2004). In this study, the researcher attempts to give meaning to the things around him in the context while taking into account as many relevant points of view as possible in order to represent a complex phenomenon.

Based on this premise, in this study, the validity of the IELTS AWM is regarded as the product of the interaction of the social context in which different social actors are taking

part for the sake of constituting social reality. In other words, it is the participants' views of the test (i.e. candidates, markers and lecturers) which are collected through different research methods, and their negotiated meanings of actions which can altogether help in building up knowledge as well as making sense of that particular social context.

### **4.3 Epistemology**

The epistemology that underpins the methodology of this interpretative paradigm is constructionism (Crotty, 1998). Constructionism, as Crotty defined it, "is the view that all knowledge, and therefore all meaningful reality as such, is contingent upon human beings and their world, and developed and transmitted within an essentially social context" (ibid, p.42). Meanings and human beings seem to be the key words in this definition. Simply, this definition suggests that human beings construct meanings through their interactions with the world they live in as a result of a sense-making process. This reveals the standpoint of constructionism as it claims that meaning is not discovered but constructed through human interactions in which objectivity and subjectivity should be held together indissolubly (Crotty, 1998). That is to say, the meanings that we hold about the world which are subjective in nature can be ultimately assisted in the construction process by the world and its objects. It is therefore important to know that the object, though meaningless in itself, can play a crucial role in the production of meaning. In addition, culture is part of this process as it shapes the way meanings are constructed.

Social constructionists have always been interested in the meaning making activities of groups and individuals around certain phenomena because meaning-making activities end up shaping the action or inaction towards a particular issue (Denzin and Lincoln, 2005).

This view, that meaning-making shapes action or inaction, is not different from Schütz's (1962) when he claimed that facts become relevant through their meanings and interpretations. For him "[a]ll facts are from the outset selected from a universal context by the activities of our mind" (ibid, p.5). Pring (2004, p.103) contended that social constructions are embedded in language and "social constructions, though maintained by social agreement, are an inherited feature of the world we are born into". Crotty (1998) pointed out that meanings in constructionism are at once objective and subjective, for example "meanings emerge from the students' interaction with the 'poem' and relate to it essentially" (p.48). He argued that "[b]ringing objectivity and subjectivity together and holding them together throughout the process is hardly characteristic of qualitative research today" (op. cit.).

In this exploratory study, the participants' views, experiences as well as their own unique way of making sense of their lives and the world around them, form the focus of the enquiry. That is to say, participants, along with the researcher who is involved in the data collection to elicit information from them, play a vital role in the construction of knowledge. Actually, the researcher has to interpret not one reality but multiple realities as knowledge is constructed from the various perspectives of the participants (Lincoln and Guba, 1985). The interest here is to understand how these realities that were constructed about a certain phenomenon have occurred. "Constructivists prefer interpretations given by the subjects or respondents themselves" such as "reflections, perceptions and stated beliefs of the actors themselves" (Druckman, 2005, p.5). Equally, negotiations of meanings and interpretations of participants' viewpoints are central to the understanding of the process. For instance, participants' responses in questionnaires can

be further interpreted in interviews. The ultimate aim behind such a confirmatory approach through verification of knowledge is to achieve trustworthiness in a rich, deep and complex constructionist tradition (Lincoln and Guba, 1985).

It should be noted, with respect to the epistemological stance, that this study follows a social constructivist view of learning where social interactions among individuals or learners play a vital role in the construction of meaning (through shared and negotiated understanding) and personal identity, rather than a radical constructivist view of learning where learners construct their own understanding (Ernest, 1994). For Gergan (1994), in stating the assumptions of social constructionism, “[t]he terms and forms by which we achieve understanding of the world and ourselves are social artefacts, products of historically and culturally situated interchanges among people” (p.49). Based on this assumption, the epistemology of this study is concerned with the construction of knowledge as it is carried out in processes of social interchange where language plays a significant role in such processes.

#### **4.4 Rationale for choosing the qualitative approach**

Both qualitative as well as quantitative methods are concerned with studying phenomena (Libarkin and Kurdziel, 2002). However, in qualitative research the researcher seeks depth rather than the breadth of coverage which is sought in quantitative research (Ambert et al., 1995). According to Cresswell (1994, p.1) a qualitative study is defined as “an inquiry process of understanding a social or human problem, based on building a complex, holistic picture, formed with words, reporting detailed views of informants, and conducted in a natural setting”. That is to say, a small group of people is researched in depth to get as much information as possible from the participants instead of taking a

large representative sample. Most of the questions in qualitative research are open-ended. This, of course, gives the researcher a chance to examine the participants closely and thus gain much more detail than in quantitative research. After obtaining these details, the researcher builds theories and hypotheses from them. McRoy (no date) claimed that “[h]ypotheses emerge from the observation and interpretation of human behaviour, leading to further observations and the generation of new hypotheses for exploration”. This means that the process of qualitative research is inductive.

Qualitative research looks closely at what the participants do and say in their natural settings, attempting to make sense of, or interpret phenomena in terms of the meanings people bring to them (Denzin and Lincoln, 2005). In this regard, qualitative research is concerned with non-statistical methods of inquiry and analysis of social phenomena. Kaplan and Maxwell (1994) have asserted that the main aim behind understanding a phenomenon from the point of view of the participants and its particular social and institutional context is largely lost when textual data are quantified. That is because the purpose of qualitative research is to help the researchers understand their participants clearly as well as the social and cultural contexts within which they live. “Qualitative researchers deploy a wide range of interconnected interpretive methods, always seeking better ways to make more understandable the worlds of experience they have studied” (Denzin and Lincoln, 2005, p.21).

The data for this study will be mainly qualitative as opposed to quantitative. Although quantitative data are quicker to collect and less time-consuming to analyse, positivist researchers adopting quantitative methods give little reference to the meaning of the observations to the subject of investigation as they view events from the outside

(Bryman, 1984) and thus produce knowledge which may be too abstract and general to be applied directly to specific contexts and individuals (Johnson and Onwuegbuzie, 2004). Therefore, an insider's view through the eyes of the interpretivist researcher will be applied in this study to provide rich details about the context of the study where findings are intended to be unique to the participants and less generalisable to other settings. Throughout the research process of this study, the researcher's aim is to interpret and evaluate every aspect of the given phenomenon in context in order to see what assumptions about reality have been made and if they are being channelled through an appropriate, rigorous and systematic enquiry. In doing so, the researcher in this study becomes involved in an interpretive process where he goes iteratively through a hermeneutic circle looking deeper and deeper to gain an understanding of the phenomenon. Fortunately, this interpretive process can sometimes help the researcher to unearth some of the meanings that even participants or authors have not considered. Crotty (1998, p.91) pointed out that "interpreters may end up with an explicit awareness of meanings, and especially assumptions, that the authors themselves would have been unable to articulate".

#### **4.4 Researcher's role**

To gain an understanding of the phenomena of inquiry, qualitative researchers adopt a unique position through the 'human-as-instrument' concept for data collection and analysis (Lincoln and Guba, 1985). In this position, the researcher basically adapts him/herself to the changing situation to capture the complexity of the situation under investigation for the purpose of data collection and analysis. Actually, this flexible position taken by qualitative researchers throughout the research process is what perhaps

makes them inevitably bring their own personal backgrounds, for example, to each interview that shape the encounter, and therefore affect the nature of data collected (Green, 2007). In addressing this problem, he suggested that an outsider who may not understand responses that assume a level of insider knowledge should share the findings with the interviewees and ask them to check their accuracy, whereas an insider who fails to attend to interview topics because of sharing assumptions with the interviewees should address this issue through triangulation of her/his perspectives (ibid).

In this study, I am taking both of these positions with my participants as I am free to choose the desired degree of participation. First, I am considered as an insider with the candidates who are the main participants of the study. That is because we share the same language (Arabic), religion (Islam), culture (Saudi), identity (Saudis), belief, and educational background (students who are going to study abroad and taking the IELTS test). In addition, these students were travelling to a context (ESL context) of which I am already a member. As an international PhD student, I was oriented to an unfamiliar academic style of writing. Despite the fact that I studied and taught academic writing at university in Saudi Arabia, in an EFL context, it was different from the way it was taught in ESL contexts. At the beginning, I found myself an outsider but as I started my course I had to adjust myself to the new and different styles of writing, academic discourses, and research practices. This gradual adaptation into the ESL context granted me membership and hence I became an insider.

I am, therefore, researching the familiar because these students were going through what I had already gone through before. In this respect, my involvement through the similarities that I share with the candidates and the membership that I obtained facilitated access and

rapport. I had the opportunity to research my own people (through questionnaires, observations and interviews) and experience life from the same perspective. As a result, this helped me to go deeper into exploring the phenomena and gain an understanding while interpreting the data of the ways in which students made sense of the test and the relationship between the test tasks and the university tasks.

However, sharing such similarities with the students through the insider position may have affected the data collection and analysis. During the data collection process, the students asked me about key successful strategies to pass the test. I had to be careful in providing such information to them as my knowledge about the test which I gained throughout the study may affect the phenomenon that I sought to investigate. It was difficult for me to decide which information should be offered to them at that stage. In addition, adopting such a position created two roles for me: my role as a researcher and my role as a member of my community. At times, it was difficult for me to disconnect myself from the phenomenon as it was blended with my private life as a student who took the IELTS test previously. Moreover, I sometimes felt that there was a tension between these two roles as I was attached to one more than the other. But this feeling changed as I left the field and started the data analysis. The tendency, which was lasting only for a short while, to favour my role as a member of my community over my role as a researcher could be because of the deep relationships that I forged with my participants which were managed and maintained through a balance via a research agenda as suggested by Adler and Adler (1987). Furthermore, being familiar with the situation may have hindered me from noticing significant aspects related to the phenomenon under investigation. Failing to recognise these aspects may engender the problem of

superficiality of the research for being less informative. In order to avoid this problem, I used a research agenda which played a crucial role in assisting me in capturing the salient issues of the phenomenon which I sought out to explore at the outset of the study.

On the other hand, the data analysis stage could have been affected by this over-involvement and over-rapport in the field through the insider position. That is to say, the necessary analytic distance could be less than the adequate required for the inquiry and this may cause bias in the analysis. Moreover, the over-rapport which the researcher may have developed with these students could also lead the researcher to evaluate the perspectives of the other group of participants, with whom the researcher is an outsider and analyse their perspectives, through the lens of the former group (Hammersley and Atkinson, 2007). In order to bring such bias to minimum, I triangulated the data for the sake of verifying the views of the participants and to obtain trustworthy accounts about the research phenomenon. Adler et al. (1986, p.367) noted that “[a]s these data are collected, researchers employ rigorous methodological strategies to ensure the validity of their accounts, cross checking them against their common sense, their own observations, the accounts of others, and hard facts, wherever possible”.

Second, I am considered as an outsider with the IELTS marker and university lecturers. That was because I did not belong to the community of practice at their work place and I shared more differences than similarities if there were any. For instance, these differences included language, religion, culture, identity, belief, and level of education, experience and training. Specifically, the retired IELTS marker who participated in the study used to work at some IELTS test centres and she was certified and accredited by Cambridge. She had sufficient knowledge in English, i.e. 9 in the overall score of IELTS, along with other

relevant qualifications in order to be appointed as an IELTS examiner. Obviously, she came from a different culture, held different beliefs, identity and religion. Similarly, I am an outsider with the university lecturers in UK. All of them were native speakers of English and were highly qualified in research and teaching. They came from different cultures and had different beliefs, identities and religions. This difference was not just felt by me as a researcher but could also be applicable to the Saudi students who were tested by IELTS examiners or taught by university lecturers.

Given these conditions, my role with the IELTS marker as well as university lecturers became marginal. Horowitz (1986) pointed out that,

“[w]hen the personal characteristics of the researcher are distinct from those of the people being studied, and the culture of the respondents restricts the categories of persons who interact with them, researchers may be limited to a marginal position. They enter as strangers and they may find it of continuing benefit to remain marginal, negotiating peripheral evolving identities that will permit them to maximize their research opportunities” (p.427).

However, this ‘remoteness’ enabled me as a ‘stranger’ to “see patterns that may be more difficult than those immersed in the situation to see” (Simmel, 1921, cited by Collins, 1991, p.36). For example, when I asked the IELTS examiner to fill the questionnaire I noticed that she struggled when listing some of the criteria for scoring both tasks. That could be because she was scoring other tests which were similar to IELTS. Although she was not aware of that confusion, the quality of scoring was not affected because she already internalised the scoring process of each task. Likewise, when I interviewed university lecturers I discovered that some of them had high expectations of ESL

candidates writing scripts due to the IELTS test results they received. Those candidates were expected to perform like their peers who were native speakers of English. Therefore, if I were involved as an insider I would not be able to learn about these contexts. Indeed, this was also possible because the participants were interested in knowing about themselves and were unable to describe what they were doing accurately.

Although this position enabled me to obtain an objective understanding of the participants and the settings they were in, it took me long enough to gain their trust and negotiate identities with them in order to access their community of practice.

Nevertheless, I became aware of the nature of the relationship that I developed and therefore was able to successfully collect the type of data I was aiming for and avoid any problems that may arise by choosing the appropriate identity for that context.

Generally, the above discussion has shed some light on the level of participation at the research context on the part of the researcher. This discussion is essential in answering the epistemological question that Davis (1973) asked “How can we know?” which he further went on to explain it through a big philosophical question

[w]ith what eyes, thoughts, feelings, acts, assumptions and cognitions is the cultural scientist to approach and engage the human subjects of his enquiry so as to be able, in the end, to render as valid and felicitous an account of their being and doing as science, or perhaps art, will allow? (p.333).

Simply, the answer to this question is to adopt a position in research either as an insider or outsider which will determine and then affect the way the research phenomenon is approached in the data collection and analysis stage.

#### 4.5 Research design

The research design for this interpretive study is in the form of a case study (see Table 8). In this case study data was collected from the students on two occasions. The aim was to collect data from the same candidates on a number of separate occasions in time (approximately 9 months, See Appendix I for a timeframe of the research) in order to understand what happened to them across time.

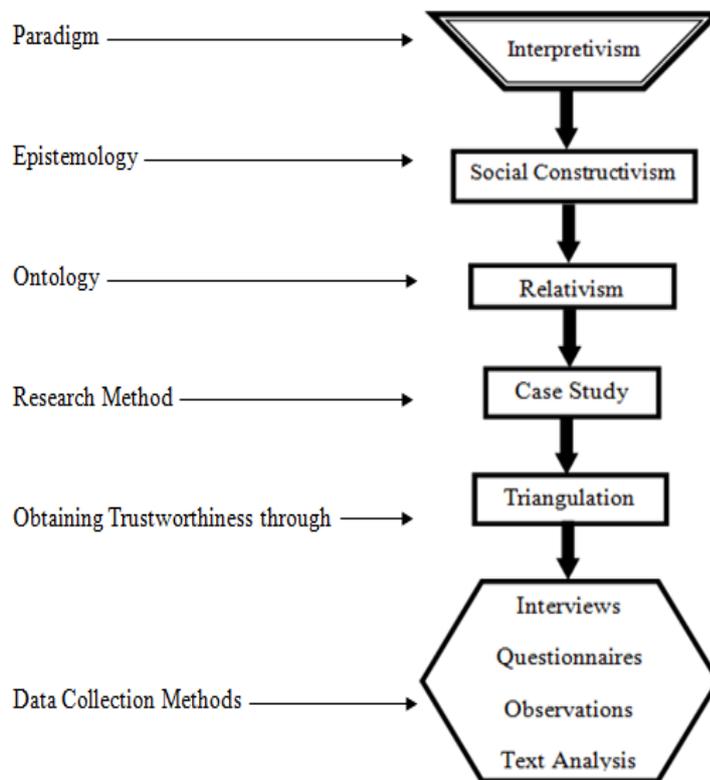


Table 8: Summary of research paradigm (ontology, epistemology & methodology), design & methods

The rationale behind using case study in this study is that it is the most suitable research method by which the researcher would be able to explore the phenomenon under investigation in its real-life context. Thus, involving the researcher in an in-depth analysis

with a detailed description and a narrow focus on the local situation provides the best approach. Yin (2009, p.4) explained that the “distinctive need for case studies arises out of the desire to understand complex social phenomena”. He defined case study as “[a]n empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” (ibid, p.18).

In this research method, the researcher is involved in the “systematic investigation of a specific instance” (Nisbet and Watt, 1984, p.74). In other words, a depth of investigation is what the case study aims to achieve through studying a part from the whole. Case studies not only “provide a unique example of real people in real situations” but also “penetrate situations in ways that are not always susceptible to numerical analysis” (Cohen et al., 2008, p.253).

In this study, the researcher made use of qualitative methods in order to identify and understand the contextual factors of the embedded and situated phenomenon through the accounts of the individuals. More specifically, in order to understand why Saudi students who met the minimum requirements of admission through the IELTS test failed to cope with the demands of the university, it is crucial to address those aspects in context which may contribute to their success, i.e. the content of *test*, the way the *students* responded to the test, the inconsistency of *markers* and the *lecturers*’ expectations of the students. In doing so, the researcher lets the emerging data speak for themselves rather than relying largely on his/her interpretation (Nisbet and Watt, 1984) which in turn may aid the understanding of the setting.

However, case studies may be biased, selective, personal, subjective, and their results may not be generalisable (ibid). Nonetheless, case studies seek to understand the complexity of people's interactions in a unique context which is difficult to unfold through statistics. In this regard, Cohen et al. (2008, p.253) argued that "there is frequently a resonance between case studies and interpretive methodologies".

#### **4.6 Sampling**

This interpretative study adopts purposive sampling as a method for selecting participants from the chosen population. Cohen et al. (2008) explained the selection process in purposive sampling by stating that "often (but by no means exclusively) a feature of qualitative research, researchers handpick the cases to be included in the sample on the basis of their judgement of their typicality or possession of the particular characteristics being sought" (p.114). That is to say, in such purposeful selection the researcher's aim in this study is to investigate the uniqueness of the case being studied.

For the researcher, participants in the study have special characteristics in common and that is why they are selected. More specifically, *candidates* in the main study were all Saudis who learned English as a Foreign Language (EFL) and taking the IELTS test to study in the UK or somewhere else where English is taught as a Second Language (ESL), the *IELTS marker* was a certified examiner for the IELTS test and *lecturers* were all teaching postgraduate international students in the UK.

This is, of course, not possible through random sampling where a representative sample is included for the sake of generalisability rather than the depth of coverage about a specific

issue. Cohen et al. (2008) argued that “[t]hough they may not be representative and their comments may not be generalisable, this is not the primary concern in such sampling; rather the concern is to acquire in-depth information from those who are in position to give it” (p.115). Purposeful sampling, as the name suggests, implies that the researcher is already aware that there are participants in the population who are probably better than others in terms of richness of information and therefore need to be included in the sample in order to provide more insight for the researcher. In doing so, “[t]he researcher actively selects the most productive sample to answer the research question” (Marshall, 1996, p.523).

This in-depth investigation through purposive sampling may be difficult to achieve if the sample is too large, due to time constraints and accessibility. Creswell (2002) has recommended the use of 3-5 participants for a case study. For fear of attrition on the part of participants, this case study suggested having 15 candidates, 5 IELTS markers and 10 lecturers in the main study in order to make sure that the number suggested by Creswell can be achieved. However, only 8 candidates, 6 lecturers and one marker agreed to continue in the study.

It is worth reminding ourselves that the main study was conducted in two settings (see Table 9), namely the College of Arts at King Faisal University in Saudi Arabia and one university in the UK. The reason for choosing more than one setting is because academic writing is not taught in Saudi Arabian universities in the way it is taught in the UK. In addition, the test measures academic writing according to the way it is taught in English speaking countries. In this sense, this study was conducted on two occasions where participants of the IELTS test were tracked, over a period of time, once they left Saudi

Arabia. After they started their academic programmes in the UK, they were contacted for data collection purposes. Written letters were given to participants and those who were in charge of the receiving institutions to obtain their permission to conduct this study. An explanation of the type of data and materials required for this study had been indicated in these letters to make sure that help can be given when needed.

<b>Participants</b>	<b>Context of the Study</b>	<b>Time &amp; Duration</b>
Candidates	Saudi Arabia & UK	Phase I: Aug. 2010- Oct. 2010 Phase II: Mar. 2011- May 2011 (Approx. 9 Months)
Marker	UK	April 2011 (1 Week)
University Lecturers	UK	May 2011 (3 Weeks)

Table 9: Participants, context of the study and time and duration

#### **4.7 Data collection methods**

The nature of the questions determines the data collection instruments. Since a single instrument cannot answer my research questions, this study employs different instruments to answer each research question. In this way, not only research questions are answered, but also triangulation is achieved. This study has made use of the following data collection methods: questionnaires, interviews and observations.

#### **4.7.1 Questionnaires**

Questionnaires in this study are used as the point of departure for other data collection methods, namely interviews and observations. Besides their wide application throughout the world and more specifically in the Saudi Arabian context they are useful instruments “for collecting survey information, providing structured, often numerical data, being able to be administered without the presence of the researcher, and often being comparatively straightforward to analyse” (Wilson, and Mclean, 1994, cited in Cohen et al., 2008, p.317). While this definition perhaps refers to the wide use of questionnaires for collecting data for large samples, it may also be applicable to small samples in specific contexts. Actually, such a difference has been associated with different types of questionnaires used. As a practical rule, Cohen et al. (2008, p.320) suggested that “the larger the size of the sample the more structured, closed and numerical the questionnaire may have to be, and the smaller the size of the sample, the less structured, more open and word-based the questionnaire may be”.

Building on this rule, this case study, due to the limited number of participants, used open-ended questionnaires in order to allow the participants to respond freely to each item in the questionnaire. Therefore, open-ended questionnaires provide more information about the respondents which otherwise might not be easily captured through closed questionnaires. Furthermore, open-ended questionnaires can “catch the authenticity, richness, depth of response, honesty and candour” which are tellingly the merits and qualities of qualitative research (Cohen et al., 2008, p.330). On the other hand, while closed questions do not allow the respondents to give their responses spontaneously, may irritate them and create bias in their responses, open questions are

time-consuming, demand more effort from the respondents and may yield irrelevant information (Oppenheim, 2000).

In this study, questionnaires (see Appendix I) were given to students in both phases to track any changes in their own perspectives (Phase I: (a) before the test in order to get background information about them and check their level of awareness about the test, (b) immediately after the test in order to get their perceptions of the test and Phase II: (c) after starting their academic programmes at university in order to see if university tasks are similar to those in the IELTS test), and marker (Phase I: the marker used to mark different samples of performance in IELTS and I wanted to know her degree of consistency and other markers when scoring the AWM in particular) (see Table 10 below) .

#### **-Construction of the questionnaires**

As mentioned above, three questionnaires were given to students and one questionnaire for the marker. The process of constructing each questionnaire is discussed in detail below.

##### **4.7.1.1 Students' questionnaires**

Construction of these questionnaires was informed by three main criteria: first, a review of related literature on language testing with respect to content validity (e.g. Messick, 1980; Hatch and Farhady, 1982; Lynch, 1982; Breen, 1985; Spolsky, 1985; Bachman, 1990; Doye, 1991; Morrow, 1991; Geisinger, 1992; Wood, 1993; Anderson et al., 1996; Bachman and Palmer, 1996; Douglas, 1997; Wall et al., 1999; Lewkowicz, 2000; Spence-Brown, 2001; Weigle, 2002; Lloyd et al., 2005; Davies, 2008) and response validity

(Henning, 1987; Bachman, 1990; Alderson et al., 1995; McNamara, 1996 & 2000); second, a review of relevant instruments designed for the same purpose or intention about content validity (Banerjee, 2000; Mickan, 2003; Mickan and Slater 2003; Green, 2007; Mayor et al., 2007; Moore and Morton, 2007; O'Loughlin and Wigglesworth, 2007; Weir et al., 2009), and response validity (Mickan et al., 2000); and third, the suitability of the instrument for the sample of the study.

In reviewing the above literature this study was better informed about the way content validity and response validity should be investigated in questionnaires. More specifically, the researcher was better informed about content validity in this study with regards to the examination of the relationship between the IELTS test tasks and the university tasks in terms of genre, length, style, level of difficulty, balance of themes, authenticity and suitability to the learners. Similarly, the researcher became acquainted in this study to the way response validity was addressed in questionnaires in aspects like meaning making where candidates reflect on the processes they went through in the test. Most importantly, the review helped the researcher to decide on which items to be included in the questionnaires that were related to the focus of investigation. In addition, it should be noted that questions about these two types of validities were integrated into each questionnaire and were not treated separately. This can be seen here in the following sections.

#### **a) Phase I**

The construction of Phase I IELTS Awareness Questionnaires (before and after taking the writing test) was mainly informed by the study of Green (2007) where he

investigated, through closed questionnaires, the students' experience of taking the IELTS test, their anxiety levels, knowledge of the AWM test format and scoring. However, questionnaires of this study were different as they were open-ended and aimed at exploring the validity of inferences of the IELTS AWM test tasks as seen by the candidates, the degree of correspondence between the IELTS AWM test tasks and university tasks, and the meaning making strategies. Nonetheless, the questionnaires were similar in terms of investigating the knowledge of the AWM test format and scoring as they may affect the candidates' performance in the test and their meaning making strategies. Pre-writing and Post-writing questionnaires are described in detail below.

#### **-Pre-writing test Questionnaire**

This questionnaire was divided into three main parts: personal information, prospective course and place of study, and specific knowledge questions about the AWM in the test. More specifically, the personal information part aimed at eliciting information about the students' age, gender, level of study, course and year of study. Similarly, the next part continued asking personal information pertaining to their prospective course and place of study, whether these students were planning to study in UK, USA, Canada, Australia or somewhere else, whether they were going to enrol in an undergraduate or a postgraduate course, and which subject they intended to study. Most importantly, the last part of the questionnaire contained a set of questions about the reason for taking the test and the overall score in the test and sub-score in writing to be achieved. It also focused on the test format and its content, i.e. the duration of the test and reporting of scores. In addition, it concentrated on the way students responded to questions, their test-taking and meaning-making strategies as well as the challenges they expected to face in the writing test.

### **-Post-writing test Questionnaire**

This questionnaire had only one main part which was specifically asking about the students' performance during the test. It investigated whether they had enough time to complete each task, whether they had a chance to brainstorm before starting to write and whether they were familiar with the tasks of the test. In addition, it also allowed the students to give their comments on the importance of guidelines/instructions in each task, the usefulness of the graph/table/chart at task 1, the level of difficulty and the frequency of going back to correct mistakes in grammar, style or organisation. Furthermore, it gave the students the chance to express their opinions about the good and bad things in the test.

### **b) Phase II**

The construction of Phase II IELTS tasks and university tasks questionnaire was informed by a review of related literature. Previous studies, reviewed here, aimed at finding correlations between the IELTS test and other related variables that may influence academic success. These variables include, for instance, language proficiency and academic outcome (Ferguson and White, 1998); IELTS scores and GPAs, academic staff ratings of student performance and students' self-ratings of performance (Cotton and Conrow, 1998); language proficiency and academic success (Dooey, 1999); the usefulness of IELTS and TOEFL as predictors of readiness for academic contexts (Hill et al., 1999); IELTS scores in reading and writing and GPA (Kerstjens and Nery's, 2000); the relationship between IELTS scores and academic performance as measured by GPA (Feast, 2002); IELTS and factors influencing academic success such as: previous professional experience, and previous English language learning experience (Woodrow,

2006); and the relationship between IELTS scores and students' 'behaviour' in a university context (Ingram and Bayliss, 2007). However, the questionnaire of this study was open-ended and more specific as it focused on the correspondence between IELTS tasks and university tasks. Such correspondence between these two entities is referred to, in the literature informing this questionnaire, as authenticity (Bachman and Palmer, 1996; Weigle, 2002; Davies, 2008). This review on authenticity helped the researcher to determine the focus of investigation in the questionnaire described below.

### **-IELTS Writing test tasks and University Tasks Questionnaire**

As mentioned above, this questionnaire was designed to examine the authenticity of IELTS writing tasks, from the viewpoint of the students, in relation to academic writing tasks that they do at university. Generally, it explored: the students' expectations of university tasks when they first sat for the IELTS test, their similarities and differences, and the suitability of the IELTS writing tasks to university students.

### **c) Pilot Study**

This section presents the findings of the pilot study which investigated some aspects of IELTS test related to content validity and response validity described above, with a focus on candidates. This section also explains the methods employed in this study and gives a brief summary of the findings and their implications for the main study. The aims were to:

- 1- Overview candidates' current test-taking practices and identify problems facing them in the in AWM in the IELTS test which are related to this study and investigate them in the main study.

2- Validate the questionnaires used for data collection.

The questionnaire was the main instrument for this pilot study. Three questionnaires (pre-writing test, post-writing test and IELTS tasks versus university tasks) were distributed to students. Selection of the students for the pilot study was based on purposive sampling as each questionnaire had different aims. It was felt that it was not practical at this stage due to the time constraints to imitate the main study in terms of following the same participants for a long period of time. Therefore, the questionnaires were given to different groups of students. However, the researcher took extra care to make sure that each group of students were taking the right questionnaires. Hence, the researcher approached two language centres in the UK because of this reason.

The director of the first language centre found that only two (i.e. post-writing test and IELTS tasks versus university tasks) out of three questionnaires were relevant because her students have already taken the test and were preparing themselves or have already started their foundation or diploma programmes. Additionally, the students who were enrolled in language courses such as pre-sessional or pre-masters had no need to sit or re-sit for the IELTS test (except if they wanted to start their course immediately) as these bridging courses could give them direct entry into university upon successful improvement in certain skills or tasks.

Instructions were given to the director of the language centre that post-writing questionnaires should be filled out by all students as they have already taken the IELTS test, whereas the IELTS tasks versus university tasks questionnaire should be only filled out by students enrolling in courses, e.g. Foundation Year, Graduate Diploma, etc., which

have similar tasks to those at university. That is because they would be able, after starting these programmes, to identify any similarities or differences between the tasks in these contexts. In contrast, students in the second language centre have not taken the IELTS test yet (but they will in future) and therefore were a suitable sample for the pre-writing test questionnaire. Students enrol in IELTS preparation courses to develop the specific techniques needed for passing the test.

### **-Development and validation of the students' questionnaires**

In order to increase the validity, reliability and practicality of the questionnaires, this study has piloted the questionnaires on a group of students (21 participants). This piloting aimed at checking issues such as coverage and format from experts and participants. Consequently, each item in every questionnaire has undergone successive cycles of development, reconstruction, validation and field-testing. In addition, this process of revising the items of the questionnaires was also based on the extensive review of pertaining literature mentioned above for each questionnaire. Furthermore, the researcher translated some parts of the questionnaires from English into Arabic to make sure that they were clear enough to the participants.

Chiefly, a review committee containing of expert researchers (1 university lecturer & 2 PhD students) was asked to check content validity and face validity of the questionnaires on the following criteria:

- 1- Identifying ambiguous, confusing or unclear words in the items.
- 2- Checking whether the item is measuring the construct that it intends to measure.
- 3- Omitting any redundant or irrelevant items.

- 4- Adding any items that are germane to the construct being measured.
- 5- Modifying sections and parts in order to move include or exclude items.

In essence, this reveals to us that anything about the questionnaire can and should be piloted (Oppenheim, 2000). After considering these criteria, the wording of some questionnaire items has been changed, some items were eliminated, and some sections were altered or created (see Appendix I for a final version of all questionnaires). For instance, there was a question in the pre-writing test questionnaire about the way the scores are reported in IELTS writing test. It caused confusion to most of the students as their replies were either not relevant, e.g., by mail or left empty with a question mark. Likewise, one expert of the review committee felt that the question about the relationship between IELTS writing tasks and university tasks is redundant as other questions are to some extent asking about the same construct, so there was no need for it. Obviously, the piloting helped to pave the way for these questionnaires by making them more lucid for the participants to grasp in the main study.

#### **d) Administration of Students' Questionnaires**

In this study, the researcher distributed three questionnaires to the students in the main study namely, pre-writing test questionnaire and post-writing questionnaire in phase I, and a third questionnaire (Students' perceptions of IELTS tasks and university tasks) in phase II of the study. More specifically, phase I questionnaires were distributed by the researcher to 15 students in the Department of Foreign Languages at King Faisal University in Saudi Arabia. However, only 8 students wanted to continue in the study. The pre-writing test questionnaire was given before taking the mock IELTS writing test

while the post-writing test questionnaire was given right after they finished the mock test. Fortunately, those 8 students showed interest in participating in phase II which took place a couple of months later. After starting their academic programmes, they were sent a reminder through emails to participate in the second phase. All of them responded to the questionnaires online through Survey Monkey website because they were far away from the researcher.

#### **4.7.1.2 Marker's Questionnaire**

Construction of this open-ended questionnaire was informed by three main criteria: first, a review of related literature on language testing with respect to scoring validity (e.g. Braddock et al., 1963; Britton et al., 1966; Odell, 1981; White, 1985; Marsh and Ireland, 1987; Hout, 1990; Hamp-Lyons, 1991b; Brown, 1991; Alderson, Clapham and Wall, 1995; Bachman and Palmer, 1996; Grabe and Kaplan, 1996; McNamara, 1996; Davies et al., 1999; Applebee, 2000; Weigle, 2002; Mayor et al., 2003; Shaw, 2002a; 2002b; 2003a; 2003b; 2004; 2005; Weir, 2005; Rignall and Furneaux, 2007; Shaw and Weir, 2007; Fulcher, 2010); second, a review of relevant instruments designed for the same purpose or intention about scoring validity (Weigle, 1994; Sakyi, 2000, Lumley, 2002; Mickan, 2003; Brown, 2006; Falvey, and Shaw, 2006); and third, the suitability of the instrument for the sample of the study.

This thorough review of literature has not only informed the study about the way scoring validity was addressed in the questionnaires of previous studies but also assisted the researcher in determining the salient aspects of scoring validity that the study should explore. In this regard, this study has narrowed its focus to checking the consistent

application of rating scales, on the part of IELTS examiners, and then identified significant features pertaining to them in the discussion below.

### **-The Questionnaire**

This questionnaire consisted of two parts: personal information and questions relating to the scoring of AWM in the IELTS test. Specifically, the first part contained 8 categories asking about the marker's age, gender, level of education, major, first language, nationality, relevant experience and training, and years of employment. The second part examined rating scales facets such as: the frequency of referring to the scoring guide during the scoring process, the importance of training sessions, the reason for awarding candidates differently in different centres or occasions, the difference between scoring Task 1 and Task 2 in terms of text features, the features of the text that markers looked for in both tasks while marking, their opinion on the validity and clarity of the scoring method, the frequency of asking for double marking, and the way they arrived to the final score. In investigating these various dimensions, which were sometimes explored but with a different skill, purpose or tool (cf. Brown, 2006; Falvey and Shaw, 2006), this study was seeking an in-depth investigation through an open-ended questionnaire and therefore was different from those reviewed above in approaching the consistency of applying the rating scales in the AWM with IELTS examiners.

### **-Administration of Marker's Questionnaire**

Due to the difficulty of obtaining IELTS markers in the study, only one IELTS marker participated in the study. Although she retired from marking IELTS papers she did not want to be known or identified. Therefore, extra care and caution was taken in order not

to cause any harm to her as a participant. She preferred to do the questionnaire online and refused to take part in the interview and the observation. When I emailed the questionnaire to her, she agreed to answer general questions only in the second part (about the test) because of the secrecy of IELTS while she left all of the first part (background information) unanswered.

#### **4.7.2 Interviews**

Cannell and Kahn (1968) defined the interview as “a two-person conversation initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him on content specified by research objectives of systematic description, prediction, or explanation” (p.527). Though this definition views the interview as conversation, it points out that it is different from everyday conversation in that a specific purpose is set out before it is conducted through a number of questions which focus on the research objectives. In addition, “it is a social encounter, not simply a site for information exchange” (Cohen et al., 2008, p.350) because in the context the “interviewers and interviewees co-construct the interview” (Walford, 2001, p.90).

In a nutshell, the interaction that is carried out between the interviewer and the interviewee constructs the interview in producing data that is placed in a social context. Baker and Johnson (1998) asserted that “interviewing may be understood as a particular form of enacting or displaying one’s knowledge of cultural forms” which helps the researcher to understand “how people make sense of each other and what resources they use to do this” (p.230). This study used a semi-structured interview where its flexibility can allow the interviewer to ask new questions during the interview while maintaining a general framework of themes to be explored. What this means is that the researcher can

understand the phenomenon in more depth than any other method. Nevertheless, such questions may sometimes be sensitive, may also make the interview “prone to subjectivity and bias on the part of the interviewer” (Cohen et al., 2008, p.352) and also time-consuming in terms of analysis.

After distributing the questionnaires, interviews (see Appendix I for a sample of each interview) were conducted with students (Phase I: (a) right after the test to reflect on the test and Phase II: (b) after starting their academic programmes at university to see if university tasks were similar to those in the IELTS test) and academics in Phase II (to know their perceptions of IELTS tasks and their expectations of students who took the test) (see Table 10 below).

### **Transcription and validation of the interviews**

In this study, the researcher transcribed the recorded interviews for approximately three months. For the sake of validation, the transcriptions were given to the participants and they were asked to make sure if their responses were reported accurately. In this process of validation, the researcher was keen to follow the suggestions pointed out by Philips (1987, p.12) when he argued that, “if the purpose of a piece of qualitative work is vivid, that is, if the intent is to give an account of how the participants in a situation see it, then checking the account with the participants is a vital step”. In doing so, the participants were actively involved in the study and the misinterpretations of their accounts on the part of the researcher were kept to a minimum.

### **Administration of Students' Interviews**

As mentioned above, the researcher conducted two interviews with the students in phase I and phase II. In phase I students were interviewed right after taking the post-test questionnaires. Likewise, phase II interviews were conducted with the students right after finishing the questionnaires. However, the method of delivery for interviews in each phase was different. In phase I interviews were conducted face-to-face while phase II interviews were conducted through telephone conversations because students were far away from the researcher. The number of students who participated in these two interviews was 8. Each student was reminded about the aim of the study before every interview. The researcher asked the students whether or not they would like to be recorded. All of them confirmed that they were happy to be recorded in both phases. In these semi-structured interviews the researcher made use of a schedule which included all interview questions. The duration for interviews was from 15mins to 25mins.

### **Administration of Lecturers' Interviews**

After finishing data collection with all of the participants, the researcher in the second phase of the study contacted 10 lecturers in two universities in the UK. All of them were invited to join the study via emails. However, only 6 lecturers agreed to participate in the study. Before starting every interview, each lecturer was given a brief introduction about the study. Also, lecturers were asked if they would like their interviews to be recorded. All of them replied positively and agreed to go ahead with the study. In these semi-structured interviews the researcher made use of a schedule which included all interview questions. The duration for interviews was from 25mins to 35mins.

### **4.7.3 Observations**

Unlike questionnaires and interviews observations are unique in the way that they allow the researcher to capture live data from the field under investigation. In addition, observations enable the researcher to “yield more valid or authentic data” (Cohen et al., 2008, p.396) because there are no discrepancies “between what people say that they have done, or will do, and what they actually did, or will do” (Robson, 2002, p.310). This may happen because people may not be aware of what they did, unable to describe it accurately or may not wish to do so. Therefore, this study made use of observational technique as it enables the researcher to observe the participants in the field and gain a deeper insight and a more accurate picture of their performance. More specifically, a semi-structured observation was used due to an agenda of issues that can be further illuminated by data gathered in a far less predetermined or systematic manner (Cohen et al., 2008). Events in the observations were documented through the use of an observation schedule (See Appendix V).

#### **Administration of observation**

In this study, observations were conducted with the students in Phase I while they were taking the mock IELTS test (see Table 10 below). The researcher observed the students in order to see how they made sense of questions while responding to test tasks. Students were observed individually. The researcher took note of the approximate time spent to understand the questions before students started writing. In addition, the researcher focused on the way students approached the tasks (e.g. planning before writing) and when exactly they started to write, pause or edit their responses.

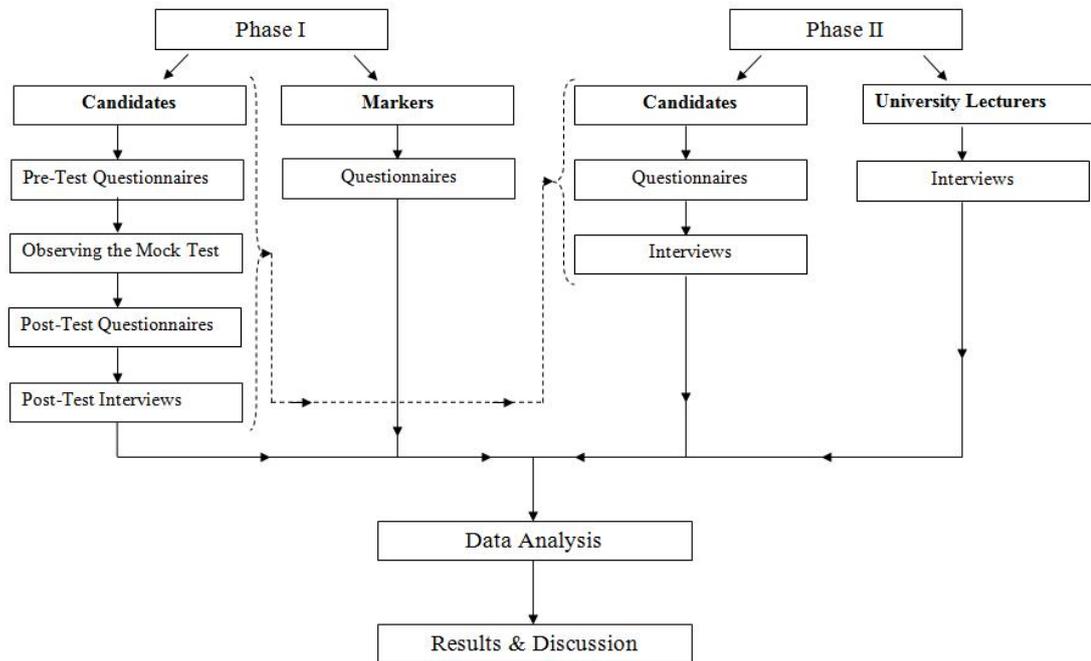


Table 10: Framework of the study

#### 4.8 Triangulation

According to Cohen et al. (2008), triangulation is “the use of two or more methods of data collection in the study of some aspect of human behaviour” (p.141). Rather than using a single method, the researcher adopts multiple or multi-methods to make meaning of the richness and complexity of human behaviour through studying it from more than one standpoint, making use of both quantitative and qualitative data (Cohen et al., 2008). In this case study, which is an example of a complex phenomenon, the researcher applies triangulation for the sake of interpreting the points of view of participants more accurately instead of relying on a single method. In obtaining relatively similar results from different methods about the phenomenon under investigation the researcher can

make sure that her/his investigation has produced, in qualitative terms, trustworthy results. Miles and Huberman (1984) argued that the purpose of triangulation is “to support a finding by showing that independent measures of it agree with it or, at least, don’t contradict it” (p.235). Denzin (1978, p.301) who is a major advocate of the use of triangulation within the interpretative paradigm listed four types of triangulation: ‘data triangulation’, ‘investigator triangulation’, ‘theory triangulation’ and ‘methodological triangulation’. This study used three kinds of triangulation, namely *data triangulation* (using multiple sources of information from different people, i.e. candidates, markers of the test and university lecturers), *theory triangulation* (by triangulating four theories about writing, i.e. writing as a cognitive practice, writing as a social construction practice, writing as a social interaction activity, and writing as a socio-cultural practice, and *methodological triangulation* (using multiple data gathering procedures, i.e. questionnaires, interviews and observations) (Denzin, 1978). These methods complement each other in an attempt to add rigour to the data being explored.

#### **4.9 Type of data**

Merriam (1998) defines data as “[b]its and pieces of information found in the environment” (p.70). It is the questions or hypotheses of the study that are composed by the researcher which determine what count as data. Therefore, one should be selective when collecting data and focus on only one aspect of a phenomenon (e.g., an event, program, etc.) because it is impossible to collect all the potential data that are related to this phenomenon (Lankshear and Knobel, 2004). They also gave the example that what counts as data when using a video recording is the selective account that is produced by the limits of the camera lens, or angle, or position of recording. Having mentioned that,

the researcher has to be careful of what to include and exclude to get the amount of data that is needed and relevant to the study. The types of data that will be used for this study are: questionnaire data, transcriptions of the interviews as well as field notes of the observations from the participants and the observer (see table 11).

<b>Questions</b>	<b>Participants</b>	<b>Methodology</b>
<p><b><u>Construct Validity</u></b></p> <p>-How do stakeholders perceive the validity of inferences of test scores in academic writing:</p>	<p>Candidates</p> <p>Lecturers</p> <p>Marker</p>	<p>Questionnaires</p> <p>Interviews</p> <p>Observations</p>
<p><b><u>Content Validity:</u></b></p> <p>1- In what ways do academic writing test tasks correspond to the types of tasks students do at university?</p>	<p>Candidates →</p> <p>Lecturers →</p>	<p>Questionnaires</p> <p>Interviews</p>
<p><b><u>Response Validity:</u></b></p> <p>2- a- What sense do students make of the tasks in the writing test?</p> <p>b- And to what extent does this affect their performance?</p>	<p>Candidates</p>	<p>Pre-Test Questionnaires</p> <p>Observations</p> <p>Post-Test Questionnaires</p> <p>Post-Test Interviews</p>
<p><b><u>Scoring Validity:</u></b></p> <p>3- To what extent are the rating scales used consistently by markers?</p>	<p>Marker →</p>	<p>Questionnaire</p>

Table 11: An overview of questions, participants and methods

#### **4.10 Data analysis**

The data analysis meant to be an on-going process in the study through an iterative interpretation of the data. Frechtling and Sharp (1997, p.4-1) indicated that “in qualitative evaluation, data collection and data analysis are not temporally discrete stages: as soon as the first pieces of data are collected, the evaluator begins the process of making sense of the information”. That is to say, the process of data analysis was done reciprocally. In this study, data were analysed inductively by following a content analysis by using themes as the unit of analysis. This grounded approach is appropriate to understand the complexity of the qualitative data. This is described in further detail below.

##### **4.10.1 Content analysis**

Weber (1990, p.15) defined content analysis as a process in which “many words of texts are classified into much fewer categories”. In other words, the researcher aims to categorise the main ideas into themes for the purpose of analysis. This coding of data makes it easy for the researcher to identify which themes are occurring frequently in the texts. Hammersley and Atkinson (2007, p.162) suggested that “[t]he initial task in analysing qualitative data is to find some concepts that help us to make sense of what is going on in the case or cases documented by the data”. This of course would require a lot of re-reading of the data on the part of the researcher in order to make sense of what is being said or done.

Data in this study were organised thematically. That is to say, themes were identified and then presented in the form of headings for ease of understanding. I coded the data to identify the themes and reconstruct them in a way that they were made meaningful to me.

This was done through ‘disassembling’ and ‘assembling’ (Charmaz, 1983), ‘sorting’ and ‘sifting’ (Jorgensen, 1989). Although these strategies for data analysis can be useful, Agar (1991) claimed that they do not give a full account of the whole data and therefore other strategies need to be added to them. Accordingly, he suggested two useful strategies. First, in order to develop their critical sense, researchers need to have “a little bit of data, massive amounts of thinking about that data, and slippery things like intuition and serendipity” (Agar, 1991, p.193). Second, with a “couple of stretches of transcript” laid on a table, researchers “need to mark different parts in different ways to find the pattern that holds the text together and ties it to whatever external frame” they are trying to construct (op. cit.). The themes in this study were coded and analysed through a computer programme called NVivo.

#### **4.11 Ethical issues**

This study recognises the importance of establishing an ethics of respect between the researcher and the participants. Its ethical considerations were derived mainly from the British Educational Research Association (BERA) guidelines. In applying such guidelines it was hoped that the anonymity and the confidentiality of the participants can be assured. For instance, names of participants or places were not disclosed but pseudonyms were used instead in order to protect their identity for fear that they may be identified or traced. Therefore, confidential information about the participants was kept as secret as possible. They were also informed that they have the right not to disclose their information. With regard to informed consent, all participants were made aware of the aims and the level of commitment in the study. More specifically, they were informed

that they had the right to withdraw from the study at any stage without giving any reason, as their participation was voluntary (see Appendix I for the consent form).

Although researchers have, in Prings' words, "the *prima facie* right to know", they must bear in mind that they have to articulate clearly their standpoint, which is bound by certain principles or guidelines for conducting research ethically (2004, p.146).

Therefore, I was keen to negotiate consent with the participants of the study. For instance, I asked the participants before each interview if they wanted to be recorded. This consent was updated on a regular basis with all participants because data collection was going to be iterative and was conducted in two settings. That is to say, in this study, data may emerge without the researcher's intention to collect them at the outset of the research. In this case, it was important, therefore, to obtain consent before starting to collect data, as well as to ensure that there was no coercion or pressure being exerted on participants to carry on with the study. Palys (1997) has drawn our attention to the importance of gaining knowledge appropriately when he contended that "[k]nowledge that rests on a foundation of manipulation, hierarchy, deception, distance, and control probably isn't worth having" (p.114).

At the end of each stage in the study, all participants were kept up to date about progress being made and initial findings that had come out of the study, through emails. This gave the participants a chance to discuss these findings with the researcher. In sharing findings with the participants, the researcher in this study can ascertain that he had done no harm to them. More specifically, the researcher can check if the participants were represented in a way they liked, and that they could take out any information that they did not intend to reveal. All confidential data in this study were kept in a safe place designated by the

researcher. For the purpose of analysis, data were entered, downloaded and then stored in a secured computer used only by the researcher.

### **3.12 Summary of the chapter**

This chapter has given a detailed description of the methodology adopted in this case study. It focused on describing the data collection methods used in the study and the philosophical assumptions underpinning the current research study. The rationale for choosing an interpretative paradigm was provided in this qualitative study as well as an explanation of the use of triangulation. The next Chapter (5) will report the results of both phases and organise them according to the questions of the study.

# CHAPTER 5

## Chapter V

### Findings

#### 5.1 Introduction

This chapter reports the findings of Phase I and Phase II of the study. These findings are all qualitative. They were gained from four *questionnaires* (the Pre-writing Test Questionnaire with the Students, the Post-writing Test Questionnaire with the Students, the IELTS Tasks and University Tasks Questionnaire with the Students, and the Marker's Questionnaire), three *interviews* (Interviews with the Students and Interviews with the Lecturers) and an *observation* with the Students. The results of this study were analysed by using NVivo (v 9.2.81.0, see Appendix VI for a sample of the analysis). The presentation of the results will be based on the research questions for this study. The order of the questions here is according to the process of research and phases of the study.

#### 5.2 Response validity question

a- What sense do students make of the tasks in the writing test?

b- And to what extent does this affect their performance?

In order to answer this question, the study utilised two open-ended questionnaires (Pre- & Post-writing Test), an observation and an interview. The aim was to check the candidates' awareness of both writing tests and know the extent to which they responded to test tasks in the way expected by test developers and the effect of such meaning making on their performance. These questionnaires together with an in-depth investigation through the

interview should at least inform us of what the test is measuring from the perspective of those candidates and hence indicate whether or not the test scores, or the inferences drawn from them, reflect the candidates' ability on the intended academic writing construct. The following sections report the findings of each tool.

### **5.2.1 Pre-writing Test Questionnaire**

This questionnaire consisted of three major parts: personal information, prospective course and place of study, and specific knowledge questions about the AWM in the test. More specifically, the personal information part aimed at eliciting information about the students' age, gender, level of study, course and year of study. Similarly, the next part continued asking personal information pertaining to their prospective course and place of study, whether these students were planning to study in UK, USA, Canada, Australia or somewhere else, whether they were going to enrol in an undergraduate or a postgraduate course, and which subject they intended to study.

Most importantly, the last part of the questionnaire contained a set of questions about the reason for taking the test and the overall score in the test and sub-score in writing to be achieved. It also focused on the test format and its content, i.e. the duration of the test and reporting of scores. In addition, it concentrated on the way students responded to questions, their test-taking and meaning-making strategies as well as the challenges they expected to face in the writing test. Results of these parts are reported in the following sections.

### **5.2.1.1 Personal information, prospective course and place of study**

Table 12 below shows personal information of participating students in Phase I and their prospective course of study. Names of students had been changed to preserve anonymity. The age range was between 24 and 36. Seven students were males with one female only. Half of the students were undergraduate and the other half were postgraduate. They were enrolled in different majors such as Applied Linguistics, English, Education, TESOL, Arabic and Accounting. Six students had finished their courses, one student withdrew from her course and another one was still continuing his current programme of study.

The majority of students (5 students) indicated that they wanted to study in the United Kingdom while two students preferred to study in Australia and only one student chose the United States. All undergraduate students were seeking to enrol in postgraduate courses which were different from their previous degrees. Of the eight prospective students only one was going to do an undergraduate course. It is clear in this table that the UK has become a place of attraction, to most of these students, for both male and female postgraduate and undergraduate candidates.

Personal Information						Prospective course & place of study		
Name	Age	Sex	Current Level	Course	Year	Place	Subject	Level
Abdu	24	M	UG	Edu	G	US	SE	PG
Abdul	24	M	UG	EN	G	Aus	AppL	PG
Ali	27	M	PG	AppL	2nd	UK	Edu	PG
Amro	26	M	UG	EN	G	Aus	AppL	PG
Essa	36	M	PG	TESOL	G	UK	-----	PG
Fatin	25	F	UG	AR	W	UK	ACC& FN	UG
Mousa	27	M	PG	ACC	G	UK	ACC	PG
Salem	29	M	PG	Edu	G	UK	Edu	PG

Table 12: Personal information, prospective course and place of study

\*PG=Postgraduate, UG=Undergraduate, G= Graduate, M=Male, F=Female, W= Withdrawn, ----- = Not indicated, UK= United kingdom, US= United States, Aus= Australia, AppL= Applied Linguistics, Edu= Education, SE= Special Education, ACC= Accounting, FN= Finance, EN= English, AR= Arabic, TESOL= Teaching English to Speakers of Other Languages

### 5.2.1.2 General and specific knowledge questions about the AWM in the test

This part consisted of 12 questions about the AWM. Students were asked 7 (Q8,Q9,Q10,Q12,Q13,Q15,Q17) general and 5 (Q11,Q14,Q16,Q18,Q19) specific questions about the AWM, and their plans throughout the test as discussed in detail below.

#### General questions

##### -Reason for taking the test

When asked about the reason for taking the test, half (4 students) of the participating students indicated explicitly that they were taking the test in order to gain entry into an academic postgraduate or undergraduate course in one of English speaking universities or institutions of higher and further education. For example,

*-“I am planning to join one of UK universities”* Essa

*-“I am going to study my masters degree in Australia”* Amro

*-“I am going to study foundation year”* Fatin

*-“I need it to join a postgraduate course”* Abdul

On the other hand, the rest of the students pointed out that they took it for other purposes.

For instance, they mentioned that,

*-“It is accepted in UK more than any test. I think it is better than TOEFL or other tests.”*  
Ali

*-“The university wants me to have it. And I want to enrol in my course ASAP.”* Salem

*-“OK, IELTS is required by a number of institutions but so as the TOEFL. I will be taking both to increase my chances for better scores.”* Abdu

It seems that they were concerned about the preferred test in the country in which they were going to study. By knowing the preferred test, they may have thought that their chances of securing a place for study were likely to increase. So, based on what Ali said for instance, the IELTS test is perceived to be better than the TOEFL or any other tests not because of high reliability or validity but simply due to its preference by target universities. These students may have guessed that from the application forms of those universities where the test was the primary choice among other indicators of English proficiency or possibly from one of their friends.

Although Abdu knew about such preference he decided to sit for both IELTS and TOEFL tests to make sure that he could meet the satisfaction criteria required by those institutions in any of these tests. This may mean that he wanted to apply to a university with the test in which he achieved higher scores, to impress a university by meeting the minimum requirements in two tests, or to check in which one he could get better scores. Thus, it could be that if he failed to achieve the required overall score in one of these tests he would not repeat the same test but rather take a different one.

Mousa was, however, having a slightly different view from the rest of the group. His main concern was to check his proficiency level in English: "*It shows if I am good or not in English*". Generally, whether the students had similar or different purposes for taking the tests, it seems that the majority were more likely to be driven by an institutional choice rather than a personal one.

*- Students' target scores and knowledge of the test*

Table 13 below shows the students' target overall scores and sub-scores by which they can guarantee entry into university. It is clear from this table that the lowest overall score is 6 and the highest is 7.5 and this applies to the sub-scores in writing too. It seems that the more demanding the course, either postgraduate or undergraduate, in terms of communicative competence, the higher the scores should be. These students were applying to different courses in different universities and countries, and there may be some sort of an agreement over the lowest scores at least among this group of students.

In addition, this table also shows the students' responses to questions about duration of the writing test and task 2, and word limit in the first task. Most students have got the answers right except for three undergraduate students, i.e. Abdu, Abdul and Fatin. Generally, inaccuracies appeared frequently in the two questions about timing, especially for task 2. Nonetheless, there was fairly high level of awareness of test content among these students. For instance, all students (as shown in the table below) confirmed that there was a difference between the 1<sup>st</sup> and the 2<sup>nd</sup> writing task in terms of: length, type, difficulty, time, word limit, topic, style, questions, and relevance (general 1<sup>st</sup> task and specific 2<sup>nd</sup> task as one thought that task 2 was more related to university study). Moreover, they indicated that type of expected response in the test should be: accurate, interesting, free from spelling mistakes, clear, well-organised, formal, coherent, focused, completed, correct, appropriate, and not necessarily related to answers in both tasks. Ultimately, this high level of awareness may have a positive impact on their performance in that it may help them to successfully apply certain test-taking strategies, e.g. time management.

<b>Name</b>	<b>Overall-score (out of 9)</b>	<b>Sub-score (out of 9)</b>	<b>Duration of the writing test (mins)</b>	<b>Word limit in Task 1 (words)</b>	<b>Allotted time for Task 2 (mins)</b>	<b>Difference between Task 1 &amp; 2</b>	<b>Answers in both Tasks should be</b>
<b>Abdu</b>	7.5	7.5	50	750	24	The relevance. Task one is more related to the aspect in general, while task 2 is in specific	Not necessarily related
<b>Abdul</b>	6.5	6.5	45	200	30	Length and type	Accurate, interesting, and with no spelling mistakes
<b>Ali</b>	7	6	60	150	40	Difficulty, different questions and time	Clear
<b>Amro</b>	6.5	6	60	150	40	The amount of both words and time as well as the title of the topic	Accurate and meet the topic
<b>Essa</b>	7	6.5	60	150	40	Number of words	Well-organised, formal and coherent
<b>Fatin</b>	6.5	6	45	150	30	Task 1 usually a chart, so it's shorter and more simple. Task 2 is the challenging part of the IELTS test	Focused and completed
<b>Mousa</b>	7	6	60	150	40	Time, style and difficulty	Correct and appropriate
<b>Salem</b>	6.5	6	60	150	40	Length and time	Formal

Table 13: Students' knowledge of the test

## Specific questions

### -Writing processes and strategies

There are four criteria on which candidates are assessed on the writing test. These include: Task Achievement (for task 1), Task Response (for task 2), Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy. In order to make sure that the students were responding to the test tasks in the way test developers expected them to be for assessment purposes, they were asked few questions about the writing processes and strategies that they would apply before and during the test.

Students' responses indicated that they would focus on different things at the beginning of the writing test. For instance, most students were concerned with coherence (organisation and linking ideas) such as Essa who wrote that he would start with "*Making an outline*", Abdu, who wanted to "*Write an introduction of the four main thoughts about the topic. Then, write about these four thoughts in details*", and Amro who preferred to "*Write the topic sentence and try to provide examples in the body*". Other students were not only concerned with coherence but also with task response (writing relevant information) such as Fatin: "*Underline the key words, prepare for the introduction, make a plan for the topic brainstorm ...*" and Salem "*Read the questions and prepare myself and organise my ideas*". Mousa was, however, worried more about task response as he wanted to "*Read the instructions and understand them*". Cohesion (use of cohesive devices) was Abdul's main concern: "*I will try to recall some conjunctions and subordinating ones that might help me to make my writing more interesting*". Unlike the

rest of the group, Ali was aiming for task achievement as he would: *“Try my best to manage my time and to apply the rules of the test”*.

While writing, most students indicated that they would pay attention to task achievement. For example, what mattered most to these students while writing was *“The time and the information given by the examiner or the test.”* Ali, *“The time limit + for the words to be spot on”*; Abdu, *“Information given especially in first task since it has to do with comparison”*; Abdul, and *“Time, instructions and organisation”*; for Mousa, who focused on both criteria, i.e. task achievement and coherence. Similarly, grammatical range and accuracy was also of concern to other students during the writing process. For instance, they cared about *“Grammar and structure”* Amro; *“Grammar rules, spelling and punctuation”* Essa; and *“Spelling and grammar mistakes”*, Salem. Coherence and cohesion were Fatin’s most important criteria while writing as he concentrated on *“The figures, the numbers, and the linking words”*.

To successfully achieve the above mentioned criteria students had to make use of certain test taking strategies for the writing test. The majority of the students reported that they considered planning and revision strategies to be the most important ones in the writing test. For instance, Fatin would *“Make a good plan, organise my ideas, follow the writing instructions and care for the spelling mistakes”*; Amro would be *“Paying attention to the body and providing a conclusion at the end of the task”*; Abdu would be more specific as he wanted to *“1- Appoint roles three main thoughts. 2- Introduction 3- Detail the main thought. 4- Strong ending”*; and Mousa who would care for *“Organisation, checking spelling and drafts”*. Perhaps, due to time constraints, some students indicated that they would make use of time management strategy together with one or both of the

aforementioned strategies. For example, Salem would apply them all: “*Time control, revision and organisation*”; Essa would use both “*Time-management and planning*”; and so did Ali as he would be concerned with “*Neat writing and time management*”. On the other hand, Abdul was not sure which strategy to use: “*I don’t have an exact strategy/ies in mind. I usually do different strategies*”.

Equally important, making sense of the question and responding to it in the manner expected by the test developers is at the heart of response validity. Students were asked about the effects of misunderstanding the question of the task on their performance. Most students indicated that they would not be able to write relevant information about the task that meets the examiner’s expectations. For instance, Abdul wrote that “*I will write something irrelevant*”; Essa reported that “*I will write something irrelevant*”; Fatin stated that “*I won’t be able to give related ideas and my task will lose power*”; Salem mentioned that “*I will write about something else different from what they want*”; Amro pointed out that “*I will not provide the answers to that the examiners want*”; and Abdu described that “*It will be hard to write on a stable basis*”. Other students indicated that they would either seek help from the examiner, which is unfortunately not permitted in the test, if they could not understand the question such as Ali who said that “*I will ask the examiner*”, or would lose marks because of misunderstanding, such as Mousa who assumed that “*I will not get a good mark*”.

Lastly, students expected to face certain challenges while taking the IELTS writing test. Almost all of the students pointed out that the topic and the time of the tasks were the most challenging issues in the test. For instance, for Abdu it was “*Most importantly time limit*”; Fatin was concerned about topics only as she would be “*Shocked by unusual*

*topics, which I don't have many words about and lose the linking between the ideas*"; Mousa was also worried about *"Difficult topics because there are topics which I don't know what to write about them especially if they are new and there is no enough information about them"*; Amro thought it was both *"The time and the topic because the topic will be new to me and may need longer time"*; Salem also added that *"The shift between different styles of writing, unfamiliar topics and the time. Because it is difficult to do two things at a time"*; Ali felt that it was the consequences of time limitations which may make things harder for him: *"I think stress and anxiety because the test has two tasks so I need to finish on time"*; and Abdul was however afraid that he may run out of words during the test *"When I reach a point in which I become unable to produce words or ideas"*. Clearly, time limitations and unfamiliar topics were issues of concern to these students in the writing test in IELTS. These aspects may, as a result, hinder the students from showing their best performance in the test in a best way and therefore we may risk having an insufficient evidence of what the students are able to accomplish in the target situation under such restrictions.

### **5.2.2 Post-writing Test Questionnaire**

Following the mock IELTS writing test, students responded to a questionnaire. This questionnaire had only one main part which was specifically asking about the students' performance during the test. It investigated whether they had enough time to complete each task, whether they had a chance to brainstorm before starting to write and whether they were familiar with the tasks of the test. In addition, it also allowed the students to give their comments on the importance of guidelines/instructions in each task, the usefulness of the graph/table/chart at task 1, the level of difficulty and the frequency of

going back to correct mistakes in grammar, style or organisation. Furthermore, it gave the students the chance to express their opinions about the good and bad things in the test. Findings of this part are reported in the following section.

### **5.2.2.1 Time allotment**

The writing test lasted for 60 minutes with 20 minutes for the first task and 40 minutes for the second task. Test developers think this is fairly adequate. However, most students reported that they needed more time to finish the test, especially in the first task. For example, Salem indicated that the time was *“Not enough for task 1. I spent 20 mins on the 1<sup>st</sup> task and I need more but the 2<sup>nd</sup> task is maybe 35 to 37 mins.”*: Abdu also indicated that task 2 is less demanding in terms of time : *“Actually, I spent less time on the second task than the first one although the second one was given double the time”*: Mousa could not stick to the time limit: *“I spent around 24 minutes on the 1<sup>st</sup> task which is not good while the 2<sup>nd</sup> task was ok but I couldn’t make revisions so I finished on time”*: Ali mentioned the likely reason for running out of time during the test: *“It could be enough for task 2 but not for task 1 because there is too much to do to understand the information. I was very worried about the time so I spent around 22 mins. on task 1 and 40 mins on task 2”*; Fatin stressed that, *“It is not much time, usually I earned 3-4 minutes before the time ended”*.

On the other hand, some students were happy with the allocated time. For instance, Amro thought that, *“The time is acceptable. The first task, I spend 18 minutes whereas in the second, I spend 32 minutes”*: Abdul also managed to stick to the time limits as there is *“Acceptable timing, I spent 20 minutes doing the first task and 25 in the second one.”*

Essa said that, *“The allotted time is OK 20 minutes on task 1 and 30 minutes on task 2.”*

Generally, these students had mixed feelings about the duration of the writing test and the first task in particular. Yet, it seems that almost all of them would agree that the allotted time for the second task was sufficient.

#### **5.2.2.2 Guidelines and Instructions**

Guidelines and instructions are important in any test as they tell candidates how the responses should be presented. Therefore, they need to be clear and easy to understand, otherwise test takers may face a double jeopardy where they have to put extra effort to make sense of them and respond to the task. Actually, students are not assessed on their ability to understand test instructions but rather on their responses to assessment tasks (Bachman and Palmer, 2010). In this study, the majority of the students considered instructions in the test to be important for the understanding of the task as *“Guidelines explain the task more to the writer because the questions are different in each version of the test”* Ali; *“They make things easy for me so I don’t need to ask for help”* Mousa; *“They are important I cant understand the question without them”* Salem; *“Crucial, as they work as facilitators”* Abdul; *“It is your keyword to understand the topic, and you will understand as much as it long”* Fatin; and, *“It is helpful but it not that enough”* Amro. Similarly, Essa thought that, *“Instructions are not enough”* and this may have led Abdu to say that they were: *“Not that important in my opinion but that also depends on the task and how significant are the guidelines”*. In short, it is clear that the students valued the importance of instructions but needed more guidance in order to focus their attention on the important aspects of the task which in turn may help them to allocate the appropriate time for each part, an issue which was always of concern to them.

### 5.2.2.3 Before the test

Getting started to write may often be the hardest thing for the students do in the test. Brainstorming could help students to overcome such difficulty. In a timed test such as the IELTS there is, however, little chance to do this. For this reason, the students were asked what they did before starting to write. Nearly all of the students brainstormed before starting to write. For example, Ali indicated that, *“I did brainstorming before each task but it was shorter for the 1<sup>st</sup> one. I tried to remember the words to describe the charts and the tables.”* Fatin would *“Make brainstorm to the words and ideas. Write any words which is linking to the topic.”* Abdul would have *“Thought about some words and ideas to begin with”*; Essa reported that he would start by *“Reading the instructions and making an outline”*; Amro also stated that, *“I write an outline to guide me through my writing”*; Salem mentioned that in the two tasks: *“I organised my ideas and made an outline in the 2<sup>nd</sup> task but for the 1<sup>st</sup> only brainstorming in my mind”*; Mousa described specifically what he did: *“In the first task I was reading the graph and the table and trying to understand them. In the second task, I was trying to think about what to write on the two sides and then give my opinion”*; while Abdu would only: *“Actually I read the requested and given materials”*. Despite the time limitations and the nature of the test, most of these students have made use of brainstorming to organise their ideas. This could in turn keep them focused on the topic and help them to write relevant information.

### 5.2.2.4 Difficulty of the questions

Most of the students indicated that the first task presented a challenge to them for a variety of reasons. For example, some students stressed that there was too much

information in the rubric which caused confusion, as Fatin reported that, *“The question in the first task was longer and asked too many things.”* Abdu also stated that, *“I don’t remember which difficult words. I remember that the first task was harder because I didn’t know what should I focus on”*; Salem was confused with words as the: *“1st task was not clear because of the word summarise (it is not an essay) and make comparisons because they ask two things at a time”*; Ali could not establish a connection between the two, i.e. the table and the chart: *“In the 1st task, there was a table and a chart and it took a long time from me to know the relationship between them. In the 2<sup>nd</sup> task the question was more clear”*; Mousa had a problem with both the table and the graph, and the words as he mentioned that

*The first task was difficult because I can’t explain the graph and the table in a 150 words and maybe the difficult words were the difference between water use and water consumption. In the second task, the question was clear but I couldn’t give examples from my experience. No difficult words.*

Like Mousa, the rest of the students found the words of the questions of both tasks to be free of any problems causing misunderstanding or confusion. For instance, Amro said that, *“It is normal questions. Nothing were so difficult”*; Essa pointed out that there were *“No difficult words”*; and Abdul indicated that, *“The questions were good, none”*.

### **5.2.2.5 Grammar mistakes, style and organisation**

When asked about the frequency of going back to correct grammar mistakes, style and organisation while writing and with which task, students had mixed responses. Some students reported that they made revisions while they were writing. For example, Fatin described that she went back while she was writing: *“About 2-3 times, and I did this in the second task often”*; Abdu pointed out that, *“I did this in the first task more but it was*

*only because of the time limit and the impeguity of requirement*". Other students indicated that they made revisions at the end. For instance, Amro mentioned that, *"I often go at the end when I review my writing and I did it with the second one"*; Ali confirmed that, *"Yes, sometimes because I had to finish within the given time so I didn't care about my mistakes at the beginning. I made a revision when I finished everything at the end"*. On the other hand, some students made revisions while writing, and at the end. For example, Salem stated that with the *"1<sup>st</sup> task I did this while I was writing. 2<sup>nd</sup> task I had time to go back. I think I did this a lot with the 2<sup>nd</sup> task"*; Mousa also explained that in the *"First task I did organisation only. In the second task I did them all many times while writing and at the end"*. However, the rest of the students had never made any revisions, such as Abdul who demonstrated that, *"Usually I don't go back to correct grammar mistakes ,etc."*; and Essa who reported that, *"I didn't"*. Actually, this variation in the students' responses could be attributed to the kind of task, its requirements, and the time limitations associated with it.

#### **5.2.2.6 Graph/chart/table in task 1**

The first task in the academic writing test in IELTS requires the students to write a description of some visual information presented in a graph, a table, a chart, or a diagram in no more than 150 words. In this study, students were given a sample IELTS writing test task where there was a graph and a table. Almost all of the students reported that having two things at a time, i.e. the graph and the table, was confusing to them and also added more difficulty to the task. For example, Salem pointed out that, *"They were confusing"*; Abdu indicated that, *"It wasn't more useful than confusing"*; Amro mentioned that, *"It is not much useful which the first one had different ideas from the*

table”; Essa suggested that, “*The first graph is enough. There is no need for the table*”;

Ali explained that, “*They have a lot of information that made it difficult for me to put them in 150 words*”; Mousa described that

*I had to look specifically for each country. I don't know what to start with is it both or one of them and which one should it be. Also, the table focused on year 2000 so there was no need to have other years. It was confusing to me I don't know if I have to use them or not. I thought I had to use everything there.*

Abdul added that, “*It useful but sometimes confusing*”. Yet, Fatin found them to be: “*Very useful, and it makes the task as copy-paste*”, but she may not be aware that she would be heavily penalised if she copied irrelevant information. Generally, this task caused some confusion to most of the students and they did not know what to focus on because information was presented in two different forms. In addition, they were asked to summarise the main features in them in a very restricted time and word limit.

#### **5.2.2.7 Familiarity with the content of the task**

The IELTS test construct, with all of its four skills, especially writing, is derived from general features of academic language use as tests cannot be authentic but what they can best do is to simulate authenticity (Davies, 2008). Accordingly, students are supposed to be familiar with most of the subjects or content of graphs, diagrams, charts, or tables presented in the academic writing tasks. However, the majority of the students in this study reported that they were less familiar with the content of the writing tasks, especially the first task. For example, Salem explained that the “*1<sup>st</sup> task has new information to me but the 2<sup>nd</sup> is something that I already know and can think of easily*”; whereas Fatin indicated that (the content of the two tasks), “*It is completely new to me*”; Mousa also pointed out that

*All of them were new to me but I knew the way I should answer them so I didn't panic because I learned some words about how to describe a graph and a table and learned some words for 2 sided arguments.*

Preparation may be helpful, as Mousa reported, to familiarise oneself with the test, as Abdu was: *“Familiar enough to write a better writing, if I knew what was really required in Task 1 only”*; after thorough preparation Essa described: *“I was ready for the test”*; or retaking the test as Amro: *“I already took an IELTS test before 10 months”*. Yet, Ali noticed that some candidates may perform better not because of preparation or retaking the test but because he thought that the test favoured them, as he argued that, *“I prepared for the test so I knew what type of things to see but I think that people from Brazil or Congo will do much better than me in the first task”*.

To sum up, familiarity with the content meant different things to these students. Some have considered it to be familiarity with the task or the question and how to answer it by following a specific genre. This could be mastered through preparation. Other students have seen familiarity as being able to know the content of the task by checking the given information in the task against their prior knowledge about the subject or the content of the task. The latter is what the IELTS test claims to measure and what caused confusion to some of the students. What those students did not know was that examiners did assume basic knowledge of the test format which could aid the understanding of the task requirements and that the test was not biased, as the test construct was based on general academic content which should be accessible to all students enrolling in higher education courses regardless of their different specialisations.

### 5.2.2.8 Advantages and disadvantages of the IELTS test

Table 14 below lists the advantages and the disadvantages of the IELTS test as suggested by the students in this study which generally concerned issues regarding test format, test preparation, test scores and the interpretation of scores, and test usefulness.

Name	Advantages of the IELTS Test	Disadvantages of the IELTS Test
<b>Abdu</b>	“I think that IELTS allow less intense time limit than the rest of the standardised tests.”	“The problem is with impeguity with sometimes with required task and focusing on the some abstract unclear orders that only show one way to the solution that way only exists in the mind of the test provider.”
<b>Abdul</b>	Not indicated	“1st task always has to do with charts & comparison.”
<b>Ali</b>	“I think because you have two tasks if you don’t make it on the first you can make it on the other and vice versa.”	“The scores given for each criteria are not known to the public. The test says they test English proficiency but I think it also tests our skills and strategies. So, we should be given marks for finishing on time or if we make outlines within this limited time.”
<b>Amro</b>	“It helps you to prepare for writing for your studies.”	“Sometimes you face general questions while sometimes you get very specific ideas which are related to medical as an example that is not related to your field of study.”
<b>Essa</b>	“Check how I am good in writing.”	“Sometimes the graphs are difficult. There is no correlation between the graphs.”
<b>Fatin</b>	“Gives an introduction to university’s assignments.”	“Time limit, spelling mistakes not accepted.”
<b>Mousa</b>	“It is a global test and many universities like it which means it is a good test.”	“It does not allow informal English although it is taken from real-life situations. I find it difficult to write about a topic I don’t know.”
<b>Salem</b>	“I think they are different each time so no one can guess the answer.”	“The time is short for the 1 <sup>st</sup> task and the questions were not clear. Also, the 2 <sup>nd</sup> task is difficult because you have to give a sound argument in a short time.”

Table 14: Advantages and Disadvantages of the IELTS test

### **5.2.3 Observations with students in the mock test**

As mentioned above, the aim of the observations in this study was to understand the way students responded to test questions, their test-taking and meaning-making strategies, as well as the challenges they faced in the writing test. Specifically, the researcher took note of the approximate time spent to understand the questions before students started writing. In addition, I focused on the way students approached the tasks (e.g. planning before writing) and when exactly they started to write, pause or edit their responses.

The researcher found that students spent different times while responding to the test tasks. Some students started writing immediately after reading the question (average time was 30-60 sec.), while the other students were stuck with the question (average time was 1-2 mins.) and did not start writing until they understood the demands of the tasks. The students occasionally referred to the prompt of the tasks to make sure that they covered every aspect of the question. In other words, they analysed the tasks and dissected them bit-by-bit. Most importantly, they were scanning for the keywords in the questions and as soon as they identified them and understood the requirements of the tasks they started writing.

It seems that the delay time while responding to test tasks may have affected the use of certain test taking strategies. For example, the researcher found in this study that the students who spent more time in making sense of the question had less time for revisions at the end of the task. Yet, some students were revising as they were writing and therefore did not have a problem with time management. Due to time restrictions, some students were seen to be planning as they were writing but this caused them to pause longer than those who planned before writing. In sum, this reveals to us that the students have

approached the tasks differently and very often their choice of test taking strategies was deliberate.

#### **5.2.4 Interviews with candidates (Phase I)**

After administering the post-test questionnaire, semi-structured interviews were then conducted with the same students. The interview (see Appendix I for the questions of the interview) followed the post-test questionnaire in order to generate more data and seek a further in-depth investigation of the students' performance during the test. Specifically, the purpose for the interview was to explore whether they had enough time to complete each task, and whether they had a chance to brainstorm or make drafts before starting to write. Moreover, the students were also asked to give their opinions on the importance of guidelines/instructions in each task, the strategies that they used while writing, the level of difficulty, task order, effort exerted, questions of the tasks, familiarity with the topic and the frequency of going back to correct mistakes in grammar, style or organisation. Furthermore, students were then invited to give their insights on the current testing practice of AWM in IELTS. Findings of this part are reported thematically in the section below.

##### **5.2.4.1 Time to make drafts and complete each task**

Making drafts often helps to improve the quality of writing and make it follow a logical pattern. This writing strategy may not be helpful enough if we know that the IELTS writing test lasts for 60 minutes where the first task takes 20 minutes and the second task requires 40 minutes. Candidates are allowed to make drafts and write on the question paper but this will not be seen by the examiner. Therefore, candidates have to write their

answers on the answer sheet due to this reason and the time limitations, as they may not have sufficient time to transfer their answers. In this study, most students reported that they were unable to make drafts, especially in the first task, and some of them could not even complete the task because of time restrictions. For instance, Abdu indicated that,

*The thing is xxx (pause) that I don't do drafts. In both tasks I never I start doing it directly but xxx (pause) in terms of the first task I did have a little bit problem in terms of the time limit. But the second one you know I finished probably before I think half or less than half the time*

It seems that Abdu was aware of the fact that he may not have enough time to make drafts and therefore he started to write as soon as he could. Although he struggled to finish the first task on time he managed to finish the second task before the cut-off time. But time alone may not be the only problem encountered by Abdu in the first task, as he mentioned that we actually need:

*more time more time but the problem is not just in the time. The main problem is that I didn't know what to write about.*

This is because he struggled with the question of the task where it said:

*Summarise the information by selecting and reporting the main features, and make comparisons where relevant.*

According to him, there was too much to do in this 20 minute short task with 150 words, including a graph and a table. Specifically, he was required to do four things: summarise, select, report and then make comparisons. It seems that the confusion was not only

caused by the aforementioned orders but also because it was not made clear to the students what to do specifically. For example, Abdu stressed that,

*say summarise ok summarise but tell me a specific thing to do*

Moreover, it was not explicit whether students should start with the graph or the table or a mix of both? Should the comparison be between Brazil and Congo or between the graph and the table? Where students were supposed to spend less time in figuring out what the question required them to do, some of them were stuck at the beginning, while they were trying to make sense of the relationship between the graph and the table. Ali, for instance, indicated that he did not have enough time in the first task:

*because I had to understand the table and the graph and then write ....you know write something related to them.*

Establishing a relationship between the graph and the table may have added more difficulty to time management. This may be because it was something implicit and was not mentioned in the question. Yet, these students have perhaps assumed that words like summarise, select, report and make a comparison would indirectly mean that there is some sort of connection between the graph and the table even if it was not declared.

The model answer that was given in the booklet started with the graph which was about global water use by sector, followed by the table which was about water consumption in Brazil and Congo. The order was deductive, i.e. from general to specific. Also, the model answer reported the performance of all sectors and then started with the table whereby the agricultural sector was selected due to its relevance to the table. Getting the answer according to this model or to the examiner's expectations within this limited time is a

fairly complex activity, bearing in mind that these students were writing in a second language that added more cognitive demands in the form of lexical resources, grammatical range and accuracy. Therefore, some students suggested that the duration of the first task should be no less than 25 minutes on average.

*25 (minutes) in the 1<sup>st</sup> task (Mousa)*

*Yes, if there is 30 minutes it would be better// (Essa)*

*Yes, I didn't make drafts. Maybe 25 minutes// (Salem)*

*Well, xxx (pause) in relating to the terms of the first task. I think we need just 10 to 15 minutes but 10 minutes will be acceptable (Amro)*

Clearly, these excerpts show that the current time duration for the writing test was another significant factor that negatively affected the students' performance.

Additionally, it could be that if this task stated that students should organise it deductively with a focus on the agricultural sector for the sake of comparisons, one would assume that the task would be clearer and students would have much more time to do the task. Based on these suggestions, it could be said that if this question was rewritten in the following way it might be more helpful to the students:

*The graph and table below give information about water use worldwide and water consumption in two different countries.*

*Report the main features of the graph followed by the table and make comparisons by contrasting the worldwide agricultural consumption of water with that in Brazil and Congo in 2000.*

This is, however, not to say that this question format is absolutely free of errors but it gives at least an indication of how the task should be handled and organised.

#### 5.2.4.2 Task guidelines

While investigating the students' concerns about time constraints in the IELTS writing test, some students spotted a problematic area in the test tasks, especially the first one. That issue was to do with the question and the need for more clarification in the form of guidelines. All students in this study stressed the importance of guidelines for various reasons. For example, some students pointed out that they work as a facilitator and make the task easier to understand:

*Yeah, sure. I think they act as a facilitator to the writer. I can xxx (pause)  
I write what you 30. asked me to do so I think it's a facilitator. (Abdul)*

*Guidelines facilitate the task to me xxx (pause) I can perform better with  
them but they should 30. be clearer in the first task. (Mousa)*

*Yes, of course. It makes the task easier we can organize our ideas by  
that// (Fatin)*

Other students, such as Amro, thought that guidelines can help those students who will be taking the test for the first time without prior preparation. He said that,

*Yes, I think it will be helpful to many students especially some students  
are not xxxxxx (long pause) you know because some students studied  
English for several years so they are taking the exam for the first time, so  
we need more xxx (pause) guidelines to help them in writing.*

It is uncommon to find students who would be coming to the test unprepared or at least know the test format and I would argue that even the carefully designed guidelines alone may not make such a big difference in performance to those students. Test preparation either in the form of self study or in a preparation course may likely give rise to appropriate cognitive processing that can enhance the students' performance in the test.

For example, if Mousa came to the test prepared he may become more familiar with the test and avoid asking questions such as:

*Interviewee: Actually, they should tell us don't write on the question paper.*

*Researcher: I think they do this every time.*

*Interviewee: But not in writing.*

Surely, it would be a waste of time to be stuck in such issues during the test as it was taken for granted that students were aware of these instructions. This was not a matter of concern for Abdul who found out that there was some significant information missing from the table in the first task. He said that,

*Researcher: Which task do you think should be clearer?*

*Interviewee: I think the first one*

*Researcher: The first one*

*Interviewee: Yes*

*Researcher: Do you want more information?*

*Interviewee: Well, personally I am not good with...xxx(pause)*

*Researcher: You mentioned that there is something missing in the graphs*

*Interviewee: Yeah, yeah. For example, in the test it was about the consumption of water between Brazil and Congo. There was some information missing for me as a Saudi student. The Congo is far away and Brazil is far away from me so I don't know the exact size of the country. So, Brazil is known as a big country but I don't know about the Congo.*

He suggested that the country size of Congo should have been given in the table as he already knew the size of Brazil. According to him, country size rather than population is an indication of water consumption. He explained that,

*Interviewee: The size depends on the consumption of water. It takes you know... it has to do with consumption of water.*

*Researcher: And if this information were given to you think that you would be able to answer the task in a better way.*

*Interviewee: Yeah, sure.*

However, the task (see Appendix II) stated in the table that water consumption on the whole was based on irrigated land and water consumption per person from the total number of the given population. This table may seem to be enough and clear to test developers and possibly there was no need to include country size as it would not be relevant or perhaps more confusing to candidates. Nonetheless, Abdul felt that the prompt was missing some significant information, namely, the country size of Congo. According to him, Congo is far away from Saudi and it was not as famous as Brazil, so he needed this information to assist him in the task. This could mean that some students imagine that the prompt would still be open to different interpretations even if choice was given. From the viewpoint of examiners who would be marking his paper, there was a misinterpretation of task prompts on the part of the student that made him/her unable to understand the purpose of the task. Unfortunately, their performance may be misjudged by the marker based on what they did.

Basically, what the students needed from the guidelines was to help them structure the task in the way that would match the markers' expectations. They believed that the current guidelines, especially in the first task, were not helpful as they left a number of things implicit such as the relationship between the graph and the table. For example, Ali and Salem advised that,

*Interviewee: Yeah, it is important ....to me xxx (pause) especially in the first task because we need to know what is the relationship between the table and the graph. (Ali)*

*Researcher: Which task you think should be clearer?*

*Interviewee: The first one.*

*Researcher: Why?*

*Interviewee: Because the table and the graph are confusing and xxxxxx (long pause)*

*Researcher: So, you wanted to know how they linked to each other.  
Interviewee: Exactly, yes. (Salem)*

This is not to suggest that guidelines should be oversimplified but test developers should make sure that task prompts are clear in order to encourage appropriate cognitive processing needed to enhance the students' performance in the test.

#### **5.2.4.3 Brainstorming**

Being aware of the importance of brainstorming in generating and organising ideas before writing, all students indicated that they were willing to perform brainstorming but only few of them managed to do so. That was because of either time limitations in the test or due to the nature of the task. For example, Abdu mentioned that,

*Interviewee: I think I was supposed to do that in the second one, I was given time I was supposed to do but I didn't do it. I'll tell you why.*

*Researcher: Why?*

*Interviewee: In the first one, I was supposed to do it but I didn't do it. In the second one, I was supposed to do it but I didn't do it for two different reasons. In the first one, I didn't have time.*

*Researcher: Ok.*

*Interviewee: In the second one//*

*Researcher: // you were supposed to do it*

*Interviewee: I was supposed to do it because you know one of the test's model way of answering any test you have to do the drafts and everything. I was supposed to do it in general but I didn't do it although I had time because it was too easy for me. I just know what to write about and therefore I started writing directly.*

Although Abdu mentioned that he would have preferred to approach the writing task with brainstorming, he did not do it. That was because he did not have enough time in the first task while the second task was too easy and perhaps familiar to him that ideas were generated when he was writing. Unlike Abdu, some students, however, described that the

second task required them to do brainstorming. For example, Essa pointed out that brainstorming was important in the second task because you had to give your opinion.

*Interviewee: In the second task yes xxx (pause) but in the first task there is no time to think xxx (pause) because your opinion is not important. What's important is only to describe what you see.*

Similarly, Ali found that brainstorming would help him in making a plan for each task, especially the second one.

*Interviewee: Yes, for each task because xxxxxx (long pause) you need to make a plan//*

*Researcher: //And did you do this in one task more than the other?*

*Interviewee: Of course. That was xxx (pause) was with the second one.*

Amro agreed with Ali that the second task required brainstorming in the form of an outline which would assist him throughout the writing process. He said that,

*Interviewee: Of course, yeah, especially for the second task xxx (pause) how to get to write briefly an outline which will help me or guide me through my writing ....*

Most importantly, Fatin explained that she brainstormed in the second task because it was more demanding in terms of thinking, generation of ideas, cohesion and coherence. She said that this was because the information was already there in the first task and all one had to do was just to rewrite them in an acceptable way.

*Interviewee: Yes, of course especially in the second task. The brainstorming is really very important with the first one it's just give you xxx (pause) some...you don't need brainstorming in the first one. I mean because it's just figures and you have to rewrite it.*

*Researcher: I see*

*Interviewee: But the second one xxx (pause) there is more thinking and bringing ideas and xxx(pause) linking words. (Fatin)*

Moreover, other students indicated that they brainstormed in both tasks despite the time limitations with no special attention given to one task over the other. For example, Mousa confirmed that,

*Yes in both tasks but it depends on the time.*

Based on this, Salem seemed to have managed his time and was able to brainstorm for about one or two minutes. Although he did it very quickly, he said that he benefited from it in improving his writing and stimulating new ideas. He expressed this as follows:

*Researcher: I mean did you think of what to do before you started writing?*

*Interviewee: Yeah, yeah I did xxxxxx (long pause)*

*Researcher: For how long do you think?*

*Interviewee: One or two minutes which is good.*

*Researcher: Yeah, it depends....time management really matters in every aspect in the writing task.*

*Interviewee: It was a bit fast so I didn't have to worry about time.*

*Researcher: Well done. I hope the brainstorming helped.*

*Interviewee: Certainly, I did better when I xxx (pause) I was thinking about it first in.....//*

*Researcher: // You mean at the beginning*

*Interviewee: Yeah.*

Abdul, in the same vein, brainstormed in both tasks but compared to Salem it took him longer (1- 4 minutes) to finish this process. Having done so, he had a clear purpose of what to do and sometimes ideas were flowing as he was writing. He pointed out that,

*Interviewee: I think but I don't take a lot of time to think. I think for one two or three maximum four minutes then I start to write and sometimes when I write the ideas come like this. I don't have to think about... I don't know the exact word to say//*

*Researcher: // you mean brainstorming*

*Interviewee: yeah*

In the above quotations, students stressed the importance of brainstorming in enhancing the quality of their writing. Yet, some of them were not able to apply it because of time restrictions or task type. Most of these students stressed that the second task needed more brainstorming than the first one. Other students, however, overcame these problems and brainstormed in both tasks. This clearly showed that where one strategy would seem to be helpful to somebody, it could be, however, less useful and time consuming for somebody else due to certain restrictions or requirements. Apparently, their choice of strategies was driven by their perceptions of what was rewarded in the test and where to place the emphasis. The next section will shed more light on the strategies students used while writing.

#### **5.2.4.4 Strategies**

In order to meet the requirements of the writing test tasks in a timed test such as IELTS, students perhaps need to consciously make specific plans to help them achieve the intended target goal/s easily within the given time. By plans I refer to writing strategies that involve cognitive processes responsible for the production of writing texts. In this study, all students indicated that they had intentionally used certain strategies in their writing. Actually, most of them focused on plans related to structure and organisation. For example, Essa mentioned that he made an outline at the beginning of the task to get him started:

*Just reading the questions carefully. Let's say writing general outlines and then start writing.*

Similarly, Amro used the outline but his purpose was to make it as a guide to save time:

*First we have to think deeply we have to get an outline which will guide us in writing. Also, we have to keep in mind that we have to finish exactly at the same time for us to...*

Other students were more eloquent and chose to describe their composing processes. For instance, Abdu mentioned that,

*the first one I did I did the first strategy I wrote you know to have and the main thought. They could be one two or three you have to put them in your mind to prepare them and then write an introduction and just talk you know around these things and then start talking about these thoughts in details and then end with a good ending.*

It seems that he had an idea of what he was going to write. This was like an outline in his mind that he could not write down possibly due to time limitations. Fatin also demonstrated that she was preoccupied with structure and organisation:

*Yeah, I mean with the introduction and the main body and the conclusion you have to put some time for both of sorry for all parts and xxxxxx (long pause) usually we leave the conclusion in the end of course and xxx (pause) in the introduction you try to write all I mean your ideas you have to xxx (pause) write what are you going to write in the main body.*

Yet, this general layout, given by Abdu and Fatin, may be germane to the second task only but may not be applicable to the first task which did not require a conclusion. Abdul, on the other hand, wanted to follow a structure that he thought would be valued by examiners. He wrote deductively focusing on coherence and cohesion devices.

*At the beginning I start writing general ideas then I go from the general to specific. I use some conjunctions and supporting conjunctions to make my paragraph you know more interesting to the reader.*

This deduction strategy may be relevant to both tasks. In the first task, it was in the form of a summary of the main ideas while in the second task it was considered to be the conclusion which was supported by the arguments.

The rest of the students indicated that revision was the most important strategy in the test besides other strategies such as time management, brainstorming, making notes, neatness, proofreading, and outlines. For example, Ali reported that he was concerned with revision and time management. He said that,

*Interviewee: I used revision and time management. Most people fail to do these//*

*Researcher: So, you think you applied them successfully.*

*Interviewee: I try to do my best but I think they are the most important ones.*

Similarly, Fatin found that she utilised a number of strategies in the test including brainstorming, making notes and revision in order to improve the quality of her writing. She mentioned that,

*Interviewee: //Yes, yes. After brainstorming we make notes and then when start writing we have a look at our notes and think of which of them you will use. Also, look after the linking words//*

*Researcher: //Look at*

*Interviewee: Yes, look at the linking words and take care with the conclusion also make our writing comprehensive//*

*Researcher: //Great.*

*Interviewee: Also, leave some time to review your writing and correct any mistakes or something you have missed.*

Salem, in the same vein, made use of brainstorming and revision but added time management to them as he may had to be more realistic with time. He stated that,

*Interviewee: Brainstorming, time control//*  
*Researcher: //Time management*  
*Interviewee: Yeah and xxx xxx (long pause) revising not sure about it*  
*Researcher: Yep, revision is a strategy*

Mousa, on the other hand, was worried about losing some marks because of poor writing or linguistic accuracy and therefore used neatness, proofreading and outlines. He explained that,

*Interviewee: Neatness because they may give me more marks...*  
*Researcher: What else?//*  
*Interviewee: //Maybe proofreading and outlines.*

Clearly, achieving accuracy through revision was at the heart of the strategies these students made use of during the test. So, it was important to know how they applied it. Abdu, for instance, reported that he would normally continue to write because he did not want to interrupt the flow of ideas when going back. But he would occasionally go back to correct grammar mistakes. He said: *“Well, usually when I write something I don’t go back, so I write it instantly. Sometimes, you know when I have especially grammar mistakes I try to go back sometimes”*. To him, idea generation might be constrained by revision so he would rather leave it to the end when he finished writing.

However, Fatin preferred to *“correct mistakes while writing but if I remember something I go back to it quickly to correct it and continue”*. She considered going back to correct grammar mistakes was *“not useful but we have to do it”*. This might be because of time limitations, as Mousa explained that, *“When I make a mistake I just cross it and don’t use an eraser because It will waste my time”*. He added that, *“I tried as much as I can to*

*avoid writing more because I remember they said in the book that the more you write the more mistakes you make/".* In doing so, he may have wanted not just to avoid redundancy and making grammar mistakes but to significantly increase the readability of his writing. It seemed that some students such as Salem had found that some criteria could be achieved with less time and effort and therefore he prioritised word count over grammar. He said: *"You know grammar is not as important as word count"*. Apparently, some students would be keen to first meet those criteria that would seem to be less controversial to examiners.

All of the above quotes demonstrate that students were conscious of choosing the writing strategies. That is to say, the cognitive processing activities of addressing and linking various stages during the production of writing did not happen by chance. If we remember that students had to present information and summarise the main features of the graph and the table in the first task and present a position, construct an argument and discuss abstract issues in the second task, we may become aware of the complexity of academic writing, especially for those second language candidates. Specifically, they had to write in a language different from their own and then had to look for the appropriate tools to make their writing easier, manageable and appealing to the reader. Most importantly, they needed to interpret the wording of the rubric in a limited time, where they underwent cognitively demanding processes, for a successful realisation of the writing task. The next section will show which task was more demanding for the students.

#### 5.2.4.5 Task difficulty

As mentioned earlier, there are two tasks in the IELTS writing test. In the first task, students were asked to summarise information given in the graph and the table and make comparisons where relevant. Basically, they needed to transfer information from the given input. In contrast, the second task required the students to bring information from outside the task. As the output in the first task was restricted to the given information, students' responses would be then highly predictable and one may expect that students in this study should face no difficulty in this task. However, most students reported that the first task presented a challenge to them and thus exerted more effort. For example, Essa indicated that the first task was difficult because of time limitations and the difficulty to understand the table and the graph. He said,

*Interviewee: xxx (pause) Umm on the first task. I think it is more difficult than the second task*

*Researcher: Why do you think so?*

*Interviewee: Because the xxx (pause) the time is limited. It's only 20 minutes.*

*Researcher: Okay*

*Interviewee: And sometimes the charts are difficult to xxx (pause)*

*Researcher: Analyse*

*Interviewee: The tasks are difficult to understand and sometimes you don't understand... let's say*

*Researcher: The general*

*Interviewee: The general view of the graph or chart*

Mousa added that there was a huge amount of information to be handled in the first task and therefore he had to think a lot to select the appropriate parts to be reported. He explained that,

*Interviewee: The second one... I would say. But that was easier than the first one.*

*Researcher: Why?*

*Interviewee: I had to think a lot.*

*Researcher: That's strange. I xxx (pause) I think most people say that it is the second one where they have to think more and bring their own information while in the first one everything is there, don't you agree with me?*

*Interviewee: Yeah, people are different but I think the reason is that because there is a lot of information you have to think how to organize them.*

*Researcher: Fair enough.*

While Mousa found that there was so much information in the task, Salem pointed out that he needed more information for the sake of clarity. He stated that,

*Interviewee: As I told you the first one.*

*Researcher: Why?*

*Interviewee: It's because it was confusing and not clear.*

*Researcher: Not clear in what sense.*

*Interviewee: In what the answer should be like*

*Researcher: It's a summary and a comparison as far as I remember.*

*Interviewee: That's right but there is little information there.*

*Researcher: Oh, really.*

*Interviewee: That's what I think.*

*Researcher: I thought you said there was a lot to do there.*

*Interviewee: Yes, a lot to do but little information*

It seems that Salem was referring to the rubric. He probably wanted it to be a step-by-step guide like those detailed task questions that he used to do in class. Also, it could be that his definition of a summary is one which he was often taught in class. It may be that a summary to him would be necessarily linked with long paragraphs or a passage where he may have a lot of information. So, the current table and graph would likely seem to him to be already summarised and that there was little information to get from them.

Ali reported that there was so much pressure on him while answering the first task and that was why he found it challenging. He said:

*Interviewee: In terms of stress the first one//*

This pressure could be because of time limitations as Abdu ran out of time and confirmed that,

*Interviewee: The first one of course.*

*Researcher: I can tell that.*

Abdul found that there was some information missing from the question such as the country size of Congo. He explained that he knew the size of Brazil but not the Congo as it was far away from him. So, it should have been given in the question otherwise the task would be biased. That is because, according to him, some students who knew the size of Congo may perform better as the task would become easier to them and thus could have a bigger chance in getting the answer right. He described it as follows:

*Interviewee: I think the first task.*

*Researcher: Why do you think so?*

*Interviewee: Because of the chance and the missing information.*

Despite the fact that IELTS test developers designed the first writing task in a way that did not require candidates to bring outside information, some candidates would still however think that some questions would be answered easily by those candidates who had prior knowledge of something. If this were true, the test would be unfair and not accessible to all students as it advantaged some students over the others.

On the other hand, other students indicated that the second task was more challenging than the first one. Ali pointed out that doing two things at a time, i.e. discussing both sides of the arguments and then giving his opinion, was very challenging to him.

*Researcher: More difficult. But why?*

*Interviewee: It could be because you have to discuss both of the views and then give my opinion.*

*Researcher: That's what's unique in the 2<sup>nd</sup> task you have to write what you think unlike the 1<sup>st</sup> one.*

Similarly, Amro added that the second task was more difficult because information was not there and while test takers needed to discuss both views and give their opinions they also had to support their ideas by evidence and examples. He stated that,

*Interviewee: In the second task*

*Researcher: And why do you think so*

*Interviewee: As I said, for the first task it's clearly the information are all available in two charts and you have to compare between them while the second one you have to discuss two different views and then give your opinions via the examples*

As a result, Fatin thought that writing information from outside the task would require more cognitive effort from the candidate, especially if we knew that the second task carried more weight in marking than the first task and more words. These two aspects were likely to put more pressure on the students. She said:

*Interviewee: The second one*

*Researcher: Why?*

*Interviewee: Because it has more words and more marks than the first one so it's quite difficult and it's the important one. The first one is easier you don't have to put your energy in it.*

To conclude, the first task was seen to present more challenge to most of these participants due to either time limitations, stress, difficulty in understanding the graph and the table, too much information in the graph and the table versus little information in the rubric, or missing information. On the other hand, the second task was also

demanding for the rest of the students in terms of bearing more marks and words, doing two things at the same time, and bringing outside information.

#### **5.2.4.6 Task order**

Upon hearing the above concerns from the students, they were asked if the task order would make any difference in terms of the task difficulty. They were generally happy with the current order. For instance, Abdul remarked that it would not make any difference: *“I think it does not matter”*. Mousa also commented that, *“this order is fine”*. In order to give more explanation for her choice, Fatin described that, *“the first one gives xxx (pause) it’s like an introduction to the other one”*. Ali recognised the importance of this order and predicted that it may affect his performance negatively if he had had started with the second task: *“I think if I started with the second one I will be tired and then may not be able to xxx xxx (long pause)//”*.

For this reason, Essa said that he was advised not to spend much effort on the first task and considered it like a warm-up for the second task, as he emphasised that he was told: *“Don’t pay much attention to the 1<sup>st</sup> task. The second task is I think more important”*. In contrast, Salem thought that his performance would have been better if he had started with the second task: *“I think if I started with the easiest//....Yeah, that would be better”*. Rather than reversing the order to start with the second one which was easier to Abdu, he suggested that, *“it’s really very important to make the first one easy”* because some students may sound worried at the beginning of the test and that may have a negative impact on their performance.

#### 5.2.4.7 The questions

As mentioned above, questions of the IELTS writing tasks can be a source of difficulty or confusion to candidates. Therefore, it is important for candidates to make sure that they interpret the questions appropriately. That is because each task has different expectations and it very often happens that the same task has different expectations in different versions of the test. Specifically, this can be seen in the keywords of the question which require the writer to respond to the task in a particular way. For example, in task one candidates may be asked to describe an object or event, the stages of a process, explain data, summarise information, or make a comparison, whereas in task two they are asked to write an essay in which they argue a case, solve a problem or respond to a point of view (IELTS Handbook, 2007). For this reason, students have indicated that they were keen to make sense of the questions. For instance, Abdu stated that, “*every time and on I look again to try to find out what I was required to do. I did that pretty much in the first one but the second one not much...probably not more than twice*”. He read the question of the first task very frequently because the task was “*so confusing I don’t even remember what they asked me to do*”. He explained in his own words that this confusion was caused because he was asked to do so many things simultaneously, namely to summarise, select, report and understand the chart. He said:

*Interviewee: (He is reading the question of the task) The the graph and table below...summarise summerise that is a general word and by selecting and reporting reporting that is another general word. You know I learned whenever you write a scientific research, you have to use specific words you don’t say summarise ok summarise but tell me a specific thing to do. And there is another word ‘report’ the main features even the main features it did say the main elements or ....I was confused what is the significance of that chart (He was pointing to the second chart)//*

Similarly, Ali stressed that he read the questions many times but that was “*because I am afraid I forgot to do something. I have a very bad memory especially in tests*”. So, it was easier for Ali to handle the questions bit by bit as he was writing. Fatin seemed to have agreed with Ali as she was very careful when she “*read every word on it and try to read it again maybe yeah...just to make sure that I am doing the right thing*”. Basically, Ali and Fatin had re-read and then analysed the questions slowly more than once to ensure that they had not missed the point. Therefore, the problem may not be necessarily with the task, as Abdu suggested, but could be with the students who misinterpreted its guidelines. For instance, Essa pointed out that, “*//the questions are clear but before you write you have to understand the thing you are going to write about and that is why I read the questions more than once*”.

Clearly, these suggestions by the students meant that they needed to engage in an iterative process of reading and re-reading of the question for a better understanding of task requirements. However, Mousa mentioned that there was no need for him to re-read the question as he already prepared for the task. He said that, “*But you know the question is always the same the change is just in the topic*”. Actually, Mousa suggested that preparation might help in understanding the question which is a time consuming activity to those students.

Additionally, students mentioned that the question helped them in writing relevant information. For example, Amro pointed out that the question defined the scope of the task as: “*it helped me a lot especially for the first task because we have to go back to the table or the graph so we can compare between them*”. Ali speculated that the thing that might help him to be on track “*could be through xxx (pause) the main words in the*

*question/*". Likewise, Fatin identified some keywords in the question which informed her of how the task should be like and what to focus on. She said that, "Every word gives you maybe from one word xxx (pause) you know what they want from you or maybe if you miss one word ...you will miss the whole task". She added that "This key word gives them a link between the task/". It seemed that Ali and Fatin started the task by scanning the questions for the keywords assuming that the goal of the task was built into them. Generally, Salem stressed that, "Questions help you to to ... focus on what you do". The process of making sense of the questions to write relevant information was perhaps done gradually. That is to say, the more the students understood the question the less likely they were go back to it again. Abdu described that

*At the beginning, at the beginning it is pretty often because I was trying to find out before I start because I need to know what to write about. But then, I didn't do a lot in the middle because I ignored it and start writing whatever I know about.*

For this reason, it could be that Mousa at some stage in writing found that "I didn't go back to the question. I went back to the graph and table ...a lot". In sum, these students have shown that writing relevant information would require a careful analysis of the question to understand the demands of the task.

Furthermore, when students referred to the questions to make sure that they were writing relevant answers, some of them indicated that the more detailed the question the more ideas they would have. For example, Salem pointed out that, "Yes, the question is important in getting more ideas". He went back to it: "Sometimes, when I don't have enough words". Also, the question was maybe a source of inspiration to Ali who confirmed that: "Of course, because if we have nothing to write about we can go to the question". In doing so, they perhaps borrowed some words from the questions and

included them in their answers. Fatin described that she copied “*Some words....some words not the whole question maybe one or two words*”. However, Ali was cautious about copying. He explained that,

*Interviewee: I tried not to copy and paste from the question//*

*Researcher: //How?*

*Interviewee: I used different words*

*Researcher: You mean synonyms*

*Interviewee: And sometimes I copy the word if I cant remember its synonym or don't know the meaning.*

So, he paraphrased the question and borrowed some words only when it was necessary for him to do so. Similarly, Mousa expressed how the question helped him to get more ideas without copying directly from it. He said:

*Interviewee: Yes, it will help me to write more//*

*Researcher: //How?*

*Interviewee: As you said I focus on one word especially on the second task and draw xxx (pause) like a mind maps shape.*

*Researcher: Good. But you don't copy the word.*

*Interviewee: Never, because it is cheating.*

This may show that these students were scrupulous in getting the most out of the questions. Unlike the rest of the students, Amro was concerned with the accuracy and appropriacy of his ideas within the given task. He stated that the more ideas he would have the more he would get closer to the target of the task or examiner's expectations: “*yeah, if it's more detailed of course it will help me a lot to to ...write correctly and accurately more details*”. On the other hand, other students commented that detailed questions would put them at a disadvantage. For instance, Abdu noticed that providing more details to the candidates “*could be confusing sometimes. It could confuse you*”. He added that what mattered to him was not the more detailed question but, “*the more*

*clear*”, as he thought that details may be misleading sometimes. Moreover, Abdul speculated that, “*I think xxx (pause) If you give me more details xxxxxx (long pause) my ideas will be limited I think but if you make it general I think I will have more ideas*”. Likewise, Essa remarked that, “*If it is detailed xxx (pause) okay xxx (pause) then xxx (pause) you will xxx (pause) there will be limitations but if it is general then you can add let’s say more ideas//*”. It may be that Abdul and Essa referred to the topic (general or specific) of the task and not to the question per se. Briefly, some candidates have shown that some textual features in the prompts have aided them to meet two important criteria in marking academic writing tasks in IELTS, namely; Task Achievement (in Task 1) and Task Response (in Task 2).

While Task Achievement (in Task 1) criterion requires candidates to write appropriate, accurate and relevant responses based strictly on the given information, Task Response (in Task 2) criterion places much emphasis on outside information where candidates need to support their positions by evidence and examples from their own experiences (IELTS Handbook, 2007). To make sure that students in this study were aware of such difference they were asked if they have related the topic of the task to what they already knew. For example, Essa found out that the second task was about change as the following prompt stated:

*Some people prefer to spend their lives doing the same things and avoiding change. Others, however, think that change is always a good thing. Discuss both these views and give your own opinion.*

Therefore, he related this theme of change to what was happening in the region during the Arab Spring in order to support his argument. He said that, “*I in fact wrote an example*

*about what happened in Egypt nowadays the revolution and the generation of facebook and so on Ok. This is a kind of change happening to our culture or our Arab nation”.*

Similarly, Ali seemed to agree with Essa that the second task did ring a bell to him. So, it was easy for him to include examples to support his ideas about change. He expressed that,

*Interviewee: Yeah, I feel that the 2<sup>nd</sup> task was more familiar to me//*

*Researcher: //In terms of content*

*Interviewee: Yeah*

*Researcher: How come?*

*Interviewee: Probably because the media talks about change nowadays*

*Researcher: Change has come isn't this what Obama used in his campaign.*

*Interviewee: Yes*

As for all IELTS tasks which are based on a communicative construct, the topic of this task was drawn from real-life contexts and it coincided here with the topical knowledge of these two students who used the Arab Spring example to their advantage.

Apparently, the rest of the students were following suit to fulfil the requirements of the second task. For instance, Mousa indicated that he retrieved some information from his memory in the second task in order to “*write more and include my ideas in the...*”. He added that, “*But I had to make sure I am writing something related to the main idea. To be honest, I was afraid to write more because the body of the essay will be bigger and bigger*”. This candidate has shown that as he may be rewarded for getting the relevant background knowledge he may be at same time also penalised for exceeding the given word limit. Hence, a balance should be maintained between these two criteria and among the rest of the assessment criteria in the writing test if successful fulfilment of task requirements was to be achieved.

Although Abdu reported that both of the writing tasks triggered his memory about something, he found that the second task allowed so much more room for topical knowledge. He described that, “*Yeah, yeah I did for both of them but for the second task I relate to it pretty much it was so easy for me to relate to the concept to the idea. I knew what I was talking about*”. In contrast, the first task to a lesser extent should not require outside information but because Abdu was stuck with the table and the graph he tried as much as he could to write something that might be close to the correct answer. He explained that,

*Interviewee: Because I didn't know what to write about in the first task. I just throw out something that I know about and I said like yeah Brazil there is the Amazon River I and I use what I yeah, this is true because I didn't understand fully what to do. I used what I already have in my background what I do know about these. Any of the given materials I used for example I knew that in Brazil there is river called Amazon and I used that and I tried as hard as I can to relate to this to anyway to the requirements.*

Being unable to spot and accurately describe the trend in this huge amount of information displayed in the table and the graph Abdu depended much more on his own background information about Brazil to accomplish the task. It seemed that he might have wanted to show the examiner that he knew something at least about Brazil (Amazon River) which may have something to do with water consumption. In doing so, it could be that he wanted to answer the task to maximise his chances of getting a high score (through task achievement, especially word count) rather than leaving it un-attempted.

To sum up, some students have demonstrated that the test perhaps focused not only on language ability needed for survival in the target language use context but also on topical knowledge manifested in their ability to reflect on their own experiences in the second

task. This issue raised about topical knowledge, may explain why some students would consider the second task easier than the first one and more relevant to real life contexts.

#### **5.2.4.8 Students' insights on the current testing practice of AWM in IELTS**

Based on the aforementioned concerns about the test in general, students came up with various suggestions which can address some of the limitations of the AWM tasks in the IELTS test. For example, Abdu argued that the current IELTS test “*doesn't necessarily measure the applicant or the student capabilities of interaction and academic writing*”. He explained that this could be because examiners had a specific answer in mind for the given task which “*means that you need one answer xxx (pause) if I encounter such situation in real life, I would answer it but not necessarily with this answer in your mind*”. That is to say, if he were involved in a target language use context he would carry out successfully the communicative demands of that context with a number of different possible responses to one cue which could still be acceptable by those people taking part in that situation. Therefore, he suggested that if writing test tasks were to reflect real life situations in measuring communicative ability, they should be designed by test developers to account for every possible response which would be accepted in target contexts.

For this reason, students such as Mousa seemed to agree with this idea as he thought that there was a specific answer required by IELTS examiners and thus he recommended that test developers should emphasise their own desired model examples which would be highly appreciated by IELTS examiners and to give less attention to those examples written by candidates. He said:

*Interviewee: I think the IELTS people should give us a model example so that we can prepare ourselves for the exam.*

*Researcher: I think they have this in their specimen materials for preparation from Cambridge.*

*Interviewee: Yes, but we need something that they write not the students.*

In doing so, according to what he said, the test would become more transparent to him through test preparation materials and the expected response could be highly predictable.

On the other hand, Ali pointed out that he had certain concerns regarding test re-taking. He described that some students may successfully get the required overall score but they may sometimes fail to meet the minimum requirements of one or more sub-scores.

Therefore, Ali suggested that in this case students should “*repeat just the modules that we failed in not the whole test*”. For example, if an institution requires that candidates whose English is not the first language should get at least 6.5 in the overall score in IELTS with no sub-score lower than 6, and it happened that a candidate managed to get the required overall score but failed in one sub-score, e.g. the writing test, he or she according to Ali should repeat just the writing test and not the whole test. The reason for his concern was because he thought that test scores may fluctuate after repeating the test where either “*the whole result may change or just one module*”.

Actually, IELTS is not a modular test as it does not allow candidates wishing to re-take the test to sit for one or more modules separately but rather the whole test simply because it intends to gather evidence of a candidate’s English language proficiency in the form of an overall profile within a specific time frame (Cambridge ESOL, 2004). Consequently, all modules need to be taken at a given point in time in order to get an accurate picture of the candidate’s overall performance in the four skills.

Re-taking the whole test instead of sitting for one or more modules was not a problem for Amro who found the test to be very general. Therefore, he suggested that there should be specific modules for each discipline as he faced difficult questions which were not related to his field of study. He explained that,

*Interviewee: yeah, sometimes when you go to the IELTS test you face you had general questions easily you can deal with while sometimes it depends on you whether you are lucky or not. Sometimes you face difficult questions which are not related to your field of study.*

*Researcher: So, your suggestions is...*

*Interviewee: There is ... this exam should be open to all students from different fields*

*Researcher: I do agree with. The IELTS test used to be like this I mean xxx (pause) and they had different modules which covered different disciplines.*

Actually, the ELTS test the predecessor of IELTS, aimed for specific testing that Amro was talking about in this interview. The ELTS used to have six modules representing the most frequent areas that overseas candidates apply to for scholarships namely: Life Sciences, Social Sciences, Physical Sciences, Technology, Medicine and General Academic (Davies, 2008). However, this practice, i.e. specific testing, was abandoned with the introduction of IELTS after several validation projects because it was felt that the test was lacking specificity and the “problem was not that the modules were too narrow, rather they were too general” (Davies, 2008, p.39). He (ibid) gave an example of some postgraduate candidates sitting for the medicine module who complained that the module contained general information about medicine and ignored their specialist field, e.g. Neurosurgery or Psychiatry that they were practising for years. For this reason, he added that, it was suggested that IELTS should reduce specificity by focusing on a single

module that deals with general features of academic language use suitable for all students enrolling in a postgraduate course regardless of their specialisation.

Nevertheless, Amro argued that the test may not always relate to general academic knowledge. He indicated that lucky candidates will face general questions which can be answered easily while unfortunate candidates may face specific (discipline-related) questions which could be difficult to answer because they were not related to their specialisation. In other words, the IELTS test according to Amro may sometimes favour some candidates over the others because of dealing with specific issues related to the subject area of those candidates rather than referring to general academic English which would be accessible to all candidates.

### **5.3 Content validity question**

-In what ways do academic writing test tasks correspond to the types of tasks students do at university?

To answer this question, the study used an open-ended questionnaire and an interview. The purpose was to find out if there was any correspondence, i.e. authenticity, between IELTS writing test tasks and university tasks. The open-ended questionnaire as well as the interview aimed at seeking an in-depth investigation from the perspectives of the participants in terms of the authenticity and the representativeness of IELTS tasks writing tasks in relation to university tasks. The following sections report the findings of each tool.

### **5.3.1 IELTS writing test tasks and University Tasks Questionnaire**

As mentioned above, this open-ended questionnaire was designed to assess the level of correspondence between IELTS tasks and university tasks. Basically, the questionnaire has one part only and it investigated: the students' expectations of university tasks when they first sat for the IELTS test, their similarities and differences, and the suitability of the IELTS writing tasks to university students. Results of this part are reported below.

#### **5.3.1.1 Students' expectations of university tasks**

When students were asked about their expectations of the university task at the time that they first sat for the IELTS test they gave mixed responses. Some students expected the university tasks to be similar to the IELTS tasks. For example, Abdul indicated that he *“was expecting that the writing tasks would be similar to the University ones”*, but he discovered that his expectation was not true as he started his programme *“while it isn't”*. Fatin expected that these tasks would be *“At the same difficult level”*. However, other students expected the university tasks to be more difficult than IELTS tasks. For instance, Salem pointed out that university tasks would be *“Much tougher”*. Amro confirmed that they would be *“Harder than IELTS test”*. Likewise, Ali specified that *“I know that they are longer and more difficult”*.

On the other hand, some students added that university tasks would be different from IELTS tasks in terms of topics, styles, discourse mode, length, and language. For example, Abdu expected university tasks to have *“Various topics and different types of writing styles. Most of tasks are argumentative writing”*. Essa was rather disappointed of the test providing insufficient linguistic ability to deal with academic language as he

*“thought it is enough for uni but it was not I think it is good exam but not same as uni eng”*. Similarly, the test caused Mousa to have unrealistic expectations of university tasks as he *“thought the study in uni just read, lesson and short writing.”* Having seen short tasks in the test, he expected the university tasks to be short as well.

Briefly, the above quotes from the students suggest that most of them have found the university tasks to be different from IELTS tasks in a number of different areas such as level of difficulty, variety of topics, styles, and discourse mode, length, and language. The next section below will shed more light on the similarities and differences between IELTS writing tasks and university tasks.

### **5.3.1.2 Similarities & differences between IELTS writing tasks and university tasks**

Table 15 below summarises the students’ views on the similarities and differences between both tasks. Basically, the students found both tasks to be similar in several areas such as language, i.e. academic English, discourse mode, e.g. descriptive and argumentative, style, e.g. objective writing, structure, cognitive demands, e.g. analyse and write, organisation, grammar and vocabulary.

On the other hand, students indicated that university tasks were more academic (vocabulary), longer in terms of time and word count, related to one’s area of study (topics), and university writing demanded doing research which would involve reading references, and conducting surveys to collect data and materials for the study. In contrast, students found IELTS to be more general (topics & vocabulary), much easier, very structured, and shorter in terms of time and word count.

<b>Name</b>	<b>Similarities</b>	<b>Differences</b>
<b>Abdu</b>	Academic English.	The university tasks are more academic.
<b>Abdul</b>	Describing the data.	Much easier. Length needed for uni is much longer.
<b>Ali</b>	Academic and objective writing style.	IELTS tasks are based on given questions and supported by some figures and concerns, and there are wide range of issues. University tasks usually are based on the area of my study. To do university tasks I need to do some surveys or researches before I start to write, Besides, I need more reference information on university tasks. Length and depth of articles.
<b>Amro</b>	In terms of structure (not always). Also, in terms of managing argument.	IELTS is very structured your always know what you will have and what you need to write.
<b>Essa</b>	I believe that they only roughly similar in the way of presenting an argument and using an academic vocabularies	They differ from each other in many respects such as time condition, various topics some of which do not have any idea about it.
<b>Fatin</b>	On both exams you need to analyse and to write.	In IELTS tasks more general than uni.
<b>Mousa</b>	I think organisations are similar.	Less number of words written and easier in terms of issues discussed.
<b>Salem</b>	Organisation and grammar, vocab...	Words count, time and vocab.

Table 15: Summary of the students' views on the similarities and differences between IELTS tasks and University tasks.

### 5.3.1.3 Suitability of IELTS writing tasks for university students

In this section students were asked if IELTS was suitable in helping them to prepare for university tasks. Most students indicated that it was somehow suitable but may not be enough on its own if students wanted to develop those skills (e.g. referencing, critical thinking, managing long papers or assignments, etc.) needed for academic writing. For example, Ali recommended that, “*the IELTS writing tasks could be considered only the first step to the University's writing tasks. In other words, student shouldn't rely on IELTS writing tasks in order to improve his ability to write about an academic topic*”. Likewise, Essa assumed that, “*To a certain extent IELTS tasks are suitable for university students because there are many similarities*”. For these reasons, Abdul suggested that the test should be reviewed again. He said: “*I think it is suitable with some amendments*”. Specifically, Amro mentioned the type of amendment he needed when he pointed out that, “*I think it is suitable if it was relevant to their subjects*”. He wanted the test to be more specific. Moreover, Fatin thought that the test would not prepare her to university but would still be “*Suitable to test our English level*”. In contrast, Salem perhaps found the test to be unhelpful as he stressed that, “*It is not suitable I think waste of money and time uni should make a new decision about IELTS if they have got a new uni English exam*”. However, other students claimed that the IELTS would help them to prepare for university. For instance, Mousa described that the test might be “*Good for university student to know what kind of essay and structure you will meet and use*”. Similarly, Abdu remarked that it “*Can make me adapt to the structure of the UK's writing*”.

To sum up, students found that IELTS test tasks were somehow similar to university tasks but may not make them ready to fully undertake university tasks. In this case,

universities should offer more support to develop academic writing, research skills, critical thinking, etc., and should not depend on the IELTS score alone to equip students with such skills needed for academic success. The next section will allow the students' voices to be heard.

### **5.3.2 IELTS Writing test tasks and university tasks Interview**

To get deeper insights from the students about the level of correspondence between IELTS tasks and university tasks, this study utilised an interview which was conducted right after the questionnaire (i.e. IELTS Writing Test Tasks and University Tasks Questionnaire). Principally, the interview explored the following themes: the relationship between IELTS writing tasks and university tasks, the role of IELTS in preparing students for university tasks, the students' perspectives on the level of authenticity of IELTS writing tasks, and general suggestions for both tasks. Findings of this part are reported in the following section.

#### **5.3.2.1 The relationship between IELTS writing tasks and university tasks**

After starting their courses at university, students were asked about the relationship between IELTS tasks and university tasks. Most of them reported that both tasks shared some textual features but with different expectations. For example, Abdu mentioned that they were related to each other as, “*Both tasks have the same structure and xxx (pause) style and er...argument*”. But it was difficult for him to maintain the same good structure, style and coherent argument in long assignments at university. He explained that the problem was that, “*when you write a paragraph in IELTS it's easy for you to focus on structure, style and argument but if you xxx (pause) if you are writing a long assignment*

*it's going to be more difficult*". Similarly, Mousa found that, *"both of them are about academic writing"*. Yet, he viewed academic writing differently in both domains when he said that, *"University tasks were more academic than IELTS tasks//"* and this was seen in *"Structure, er academic words, xxxxxx (long pause) style, and er xxx (pause) references, and so on"*. Abdul seemed to support Mousa's argument when he pointed out that, *"university tasks are academic not general academic"*. He was referring to the point of discussion when he said that IELTS tasks were *"general"* and different *"in time and it is shorter"* but the researcher corrected him by saying that they were general academic. Actually, IELTS used a general academic construct because its predecessor ELTS used specific testing and provided different modules for each discipline but it was seen that it was not specific but rather very general (Davies, 2008).

Essa also held a similar opinion about the test from the language centre where he prepared for IELTS when: *"They said that IELTS tasks were more general and university lecturers expect to see something different xxx (pause)//"*. This difference was manifested according to him in being *"more critical not just like what we do in IELTS where you have your opinion"*. Likewise, Fatin stated that, *"I would say xxx (pause) say the same style and structure but university writing is more academic//"*. By academic she meant that, *"University tasks require critical writing and deep analysis before reaching to a conclusion"*. Salem confirmed that, *"in IELTS you say your opinion xxx (pause) in the in the whole essay while in the research you cannot just let your voice speak throughout the whole assignment...you xxx (pause) need to have other people"*. He did not know about such difference and thus did not expect university tasks to be more demanding than IELTS which according to him *"gave very little things about university tasks//"*. In fact,

Ali noticed that the nature of tasks in terms of content language, i.e. concrete versus abstract, were somehow different and that, *“both of them connect to life but IELTS is more/”* and that was *“because IELTS topics are taken from everyday language not like ... not like university assignments”*. According to Amro, these tasks differed also in *“length, time xxx (pause) and er”*, *“Maybe structure”* and *“Vocabulary as well”*. Briefly, the above different expectations may have caused the students to think that there was a weak relationship between IELTS tasks and university tasks.

### **5.3.2.2 The role of IELTS in preparing students to university tasks**

The academic writing test in IELTS claims that it assesses *“whether a candidate is ready to study or train in the medium of English at an undergraduate or postgraduate level”* (IELTS Handbook, 2007, p.2). That is to say, when candidates achieve the minimum requirements in IELTS for the receiving institution they should be ready and prepared to undertake university tasks. However, students in this study reported that this kind of preparation from IELTS was not sufficient. For example, Ali expressed that, *“IELTS I think is very general and cannot be enough”* because *“there are so many things that we do now different from IELTS”*, for instance *“we read research”*, *“we write assignments based on these research”*, *“We make references”* and, *“The language of literature review is different from methodology and findings and conclusion”*. Abdul agreed that the role of IELTS to him was not to make him ready for university tasks but *“Just to know general features of academic writing but not to help you write long assignments”*. He seemed to have known this before starting university as he decided to take an academic English course besides the IELTS preparation course because: *“academic English was more related to university assignments”*. Similarly, Essa pointed out that after getting the

required score in IELTS he *“had to take a preparation course and my government said that it is important to take it because this course will teach me more academic skills ...and more xxx (pause) writing”*.

More specifically, Mousa explained that, *“It didn’t help so much. IELTS as a test cannot be exactly like university tasks”* simply because *“I make so many drafts and it takes time from me to get the final one”* and, *“I use the internet and dictionaries sometimes to help me with writing. So, I don’t think IELTS can do this”*. Likewise, Salem believed that IELTS tasks did not help him to be critical as he said that, *“I remember when I submitted my assignment in uni my tutor said that this was not academic writing. xxx (pause) I told her this is what the IELTS test was about”*. According to him the reason was because: *“The assignment was not critical. She said that you just said your opinion and gave little space for other people to argue.”*

Although Abdu indicated that IELTS *“helped me to know the structure, style and argument and also the way academic writing should be like”*, he found that he needed to know more about academic writing. He described that, *“one tutor asked me to make a critical review and he said be critical...be critical on the whole assignment and I remember that I read some books about critical writing”* whereas *“another tutor said make a summary and don’t be too much critical at the beginning”*. Amro confirmed that IELTS *“didn’t give me enough preparation”*. The kind of preparation that he needed *“can be in different forms xxx (pause) like for example how to write research papers how to to make references and...er”* and how to do *“Critical reading and writing and university life in general”*. Due to the above concerns about IELTS, Fatin suspected that it would reflect the *“Real ability to do university tasks because they are different”*. To sum up, these

students found that they were lacking essential academic (research) skills needed for success at university and that the IELTS test offered them limited support in this regard.

### **5.3.2.3 Students' perspectives on the level of authenticity of IELTS writing tasks**

In order to get an overall impression of the level of authenticity of IELTS tasks, students were asked if they were representative of university tasks. They indicated that there was some sort of similarity between the tasks in both domains, yet IELTS tasks may not accurately mirror the content and skills needed at university. For example, Amro described in his own words that IELTS tasks focused repeatedly on certain genres and ignored other important ones that students were likely to encounter in university. He said:

*Interviewee: Maybe, I am not sure. But they don't cover other important types of writing//*

*Researcher: //But these genres are the most frequent tasks that students do at university.*

*Interviewee: Yeah, but also ...other tasks will be needed.*

Also, Abdu noticed that IELTS tasks did not fully represent the format of university tasks. He said: “*Not sure xxx (pause) not in everything you know*”. He added that “*the first task should be longer and...xxx (pause)*” “*In terms of time*”. In addition, Abdul expressed that the topics of the first task in IELTS were: “*In some ways, in some ways*” similar to university tasks. He recommended that, “*such topics especially the first one needs more academic writing than the second one//*”. Similarly, Fatin agreed that IELTS topics were too general when she said that, “*No, no the tasks are different because these tasks in IELTS came in many different topics and many different subjects but we will only study one or two subjects in the university I mean when you choose your major*”. Clearly,

she was asking for the IELTS topics to be more specific and relevant for each discipline. That is because she would be asked at university to write about topics from her discipline, i.e. English, and if IELTS claims to be representative it should cover such topics.

For these reasons, Ali found that the IELTS tasks had somehow under-represented the construct in the target context. He mentioned that,

*Interviewee: It's a difficult question... you know at university there are many things students do that the test can't cover.*

*Researcher: That's what's expected from each test. It just takes a snapshot from the candidate's performance. It tests the basic linguistic knowledge not study skills.*

*Interviewee: I agree with you but if I take the test again after going to the university I may get lower grades and still Ok at my course.*

Ali seemed to suggest that success at university did not depend solely on the test. That is because if he had the chance to repeat the test he would have got lower grades while this low performance would let him survive at university. This may have led Mousa to argue that raising the minimum requirements in overall scores may not be the solution for failure or drop out at university. He thought that universities should accept candidates with “6 in the overall. Because once you go to university you learn how to do tasks you don't need to get 7 to be better. It depends on how much you study”. Salem agreed that the test alone may not be sufficient to prepare students for university courses because: “even if you get 6.5 you need to follow academic writing rules at university and xxx (pause) and learn how to use them”. It seems that Salem meant to say that the test did not reflect the kind of tasks that students would expect to see in the target context and therefore should seek additional support in the form of preparation courses for university.

Essa commented that because of such limitations of tests where “*you will never have one thing that is perfect*”, he suggested that test users should not “*make a student’s future just you know based on just this test*”. Briefly, these students have found that IELTS tasks shared some common features of target language tasks. Nonetheless, students were doubtful of the representativeness of IELTS tasks as they may not exactly mirror the content and skills needed to undertake university tasks.

Based on the aforementioned concerns, students suggested a number of issues regarding the test. Most importantly, Fatin indicated that in order to increase the authenticity of IELTS tasks test developers “*should make the tasks look like university tasks*”. She added that there was a heavy emphasis on certain genres and types of testing: “*I mean IELTS should give us different types of writing each time and not just argumentative and a summary task all the time*”. Other genres that she called for to be added may include those relating to research, i.e. proposal, literature review, critical review, etc., as they lied at the heart of university tasks. In order for IELTS to simulate this it should integrate the reading and writing modules (like its predecessor the ELTS) in read-to-write tasks for a more authentic assessment of academic writing skills. Ali agreed that, “*they should make the test according to what you are going to do*” because: “*I did not see the first task in IELTS .....when I started my course. They said that both tasks will be relevant but this is not true*”. Likewise, Essa mentioned that he did not benefit from the first task in preparing him for university tasks. He said: “*The first task I don’t see it as related to my major//*” “*Because we don’t have tables and graphs in our assignments*”.

In addition, the level of difficulty with respect to topic was not the same because Amro noticed that, “*Sometimes it is easy and sometimes it is difficult*”. He added that he: “*took*

*the test three times and only one time the topic was familiar to me*". Fatin suggested that the issue of under-representation and level of difficulty might be overcome if IELTS could *"give us four or five tasks or topics and ask us to choose two of them"*.

Furthermore, the marks that were designated for each criterion in the academic writing tasks were unknown to students. Mousa described that,

*Interviewee: I think in writing they need to let us know how they mark it.*

*Researcher: They have put the scoring criteria for the writing test in the IELTS website. Is that what you are looking for?*

*Interviewee: No, I saw these when I prepared for the test. But I need to know how much they give each criteria.*

*Researcher: Why is that so important?*

*Interviewee: We need to know how the scores are distributed so that we can focus on the important things.*

In the same vein, the band descriptions for academic writing tasks were difficult to understand. Amro felt that *"the meaning of scores is so complicated"*. These comments by the students regarding the test have shown that more improvement was needed in areas such as format, content or scoring method.

### **5.3.3 Interviews with lecturers (Phase II)**

After collecting data with the candidates, the researcher conducted semi-structured interviews with 6 lecturers. The interview (see Appendix I for a schedule of the interview) sought an in-depth investigation of lecturers' perceptions of IELTS tasks and their expectations of students who took the test. Findings of this part are reported thematically in the following section.

### 5.3.3.1 Lecturers' perceptions of IELTS tasks

In this part of the interview, lecturers were asked whether the academic writing tasks would make students ready to start their academic programmes at university, whether the academic writing tasks were relevant to and representative of the kind of tasks that students were going to do at university, and whether the practice of raising the IELTS overall scores and sub-scores by universities would ensure that only good students would be accepted into university. Each of these sub-themes will be presented below, as follows:

#### *(a) IELTS as a pathway to university*

With respect to preparing students for university tasks, some lecturers expressed their concerns over the issue of readiness that is claimed by IELTS (IELTS Handbook, 2007) and therefore were doubtful that students would be able to cope with the demands of the university. For example, John mentioned that, *“I think it’s preparing people for a general approach to academic writing but then that has to have xxx (pause) the context of the subject”*. He added that

*The IELTS system can make sure that they are able to do the kinds of argument that they used to do but do them in English but it doesn’t necessarily mean that they can do the sort of arguments that they need to do in an English course*

Therefore, this has led universities to offer more support during the academic programmes as they according to what he said, *“recognised that an IELTS score doesn’t guarantee that an individual is xxxxxx (long pause) sort of already an expert in the field they are going to study”*. Similarly, Kate argued that the IELTS test is not enough to show evidence that students have the necessary requirements that can help them to

overcome the complexities of academic writing needed at university so, *“just because people can do a short piece of writing like this doesn’t really test that they have enough knowledge and understanding of the sort of way in which you write an academic case”*.

University courses require high thinking and critical writing skills and the problem with IELTS as she thought was that, *“it gives a very rough measure. It tells us ok they can do something but it doesn’t actually tell us how people are going to cope”*.

Furthermore, Samy pointed out that, *“This (pointing at the IELTS tasks) does not prepare for this (pointing at the thesis)”*. He added that students *“would possibly write very well here (in IELTS) but it doesn’t mean that they will do well here (in thesis)”*. He explained that university tasks at postgraduate level have different contents as *“we don’t have a single task in our programme for PhD students even for Master students that asks you do you think whether people spend time playing football or their hobbies are changing”*.

Alternatively, he suggested that such tasks *“would prepare undergraduates for year one because this is the kind of xxx (pause) activities they will be doing”*.

Sue seemed to have joined the rest of the lecturers in arguing that students would not be ready for a university study. She suggested that, *“I don’t think this would capture all the things students need to do at university”*. Therefore, it may be difficult to generalise on the students’ performance in these tasks as she stated that *“I don’t know if that is enough to give an opinion on students writing”* simply because *“It’s a bit limiting”*.

#### *(b) IELTS relevance to and representativeness of university tasks*

Another important aspect regarding the IELTS test that lecturers were asked about was its relevance to and representativeness of university tasks. Some lecturers found that these

two aspects were to some extent met by the test in certain areas. For instance, Abraham indicated that, *“Both tasks are related”*, because students *“need to have argument scales, em a lot of vocabularies”* and, *“they need to understand some basic numerical cases. In this sense, they are linked and they are important”*. Likewise, John pointed out that, *“IELTS task one is a sensible task”* because it *“requires students to have the English vocabulary and sentence structure to interpret the task around into a text”* and, *“it also encourages them to think about the presentation data”*. Also, task two in IELTS *“gives the context in which people can reveal their abilities to write argumentative texts and that’s certainly very relevant”*.

However, he argued that task one may not be relevant and representative as he was *“not sure that describing the histogram is something I would expect students to do routinely as part of engagement with research data or as part of the assignments”*. In the same vein, task two was *“a little bit artificial because it’s just asking for their opinion about a particular idea”*, while at university lecturers *“would expect people to make reference to the literature”* to support their opinions. Similarly, Kate seemed to agree with John that task one may not cater for all students as it was *“a little bit a scientific”* and, it *“may be good for someone who is gonna do a science doctorate”*. The test was limited and specific in this regard whereas language according to her has *“different registers and the registers have different fields”*. Therefore, the IELTS test according to Sue has a low level of relevance and representativeness of university tasks as its tasks were just *“an element of what they’ll have to do at university”* and, *“the topics are not obviously academic”*. Samy reminded us that the purpose of the test was to show *“somebody’s*

*proficiency in English anyway*". So, there was low level of correspondence between both tasks because IELTS *"does not tell you if they can do this (thesis)"*.

(c) Raising the IELTS scores

Lecturers in this study mentioned that their institutions raised or were going to raise the IELTS overall scores and sub-scores in writing. Most of them were not happy about this practice because it wouldn't ensure that only good students would be accepted into university. For example, John argued that, *"we can't expect scores on an IELTS test to guarantee that the student is in exactly the same place to begin their postgraduate study as a student who has done their education in the UK"*. He added that the reason for being pessimistic is: *"because a student has to acquire the ability to xxx (pause) write in the styles and within the culture of UK higher Education"* and thus he thought that, *"they are missing the point in raising it from 6.5 to 7"*. Similarly, Abraham pointed out that raising the IELTS scores *"will not solve the problem xxx (pause) but actually there is a certain level where students can start from so we should not set high scores every year"*.

Moreover, he suggested that IELTS should be *"one of the criteria for admission"*, as universities *"need to make an interview with these students"* depending on the discipline in which they will specialise.

Samy seemed to be in agreement with Abraham's suggestion when he proposed that, *"I sometimes ask for a whole chapter from a thesis which is helpful or interview"*. That is because he did not *"think 7 is that different from 6.5 in writing"* and, *"We don't have a proof if they have 7 that they will be better in writing their theses"*. In addition, he stated that the IELTS test would equip students with the technical elements (e.g coherence and

cohesion, etc.) that they need in writing but, *“There is much more academic writing at a doctoral level than the technical element of writing”*.

Kate mentioned that her colleague was totally against the idea of raising the IELTS scores. She said that, *“He said that it’s such a bad measure. It’s not a good idea”*, and then she commented that, *“part of me agrees with him but the other part of me says how ...we gonna do what else is there that helps us choose”*. She concluded that, *“the IELTS is the best we’ve got at the moment even though it’s inadequate”*. Likewise, Sue expressed that

*I don’t think it would solve it however I think it still happens because they are looking for a relatively straightforward way of ensuring that they’ve got students who are better matched to what they want and of course a less straightforward way is to devise another test.*

She stressed that one may not notice a difference in the performance of students if the difference is just half a band score but with a band score difference: *“I think you do notice if somebody has got a 6 or 7 in their writing”*.

### **5.3.3.2 Lecturers’ expectations of students who took the IELTS test**

Lecturers were asked in this part about their expectations of students who took the IELTS test, and what type of support should be provided. Each of these sub-themes will be presented below as follows:

#### *(a) Lecturer’s expectations of international students*

Most lecturers expected students to be at a level in English where they can survive at university. More specifically, Abraham indicated that, *“My expectation is that students should express themselves freely without any barrier from the language”*. Similarly, Kate stated that, *“we’re wanting people understand xxx (pause) academic writing which is very*

*different*” (from the IELTS tasks). Also, Samy expressed that, “*we assume that our students are good enough in grammar*” and so “*we assume that if you have 7 you are Ok but overall this is a fine guess but we don’t always get it right*”. Sue explained in detail that students should be able to “*write reasonably clearly without too many errors. I mean it’s got to be comprehensible and intelligible*”, “*they need to be aware of features of academic discourse*” and what they “*need to do is to read and to critically interpret what they are reading and build an argument*”. David added that, “*I have high expectations of all my students*”, and therefore “*I am willing to make certain allowances for linguistic imperfections, but I expect my students to have all work read on completion by a competent linguist*”. In contrast, John would: “*expect to provide more support for international students in order to help them to reach to that standard*” and he would be “*less fussy about the quality of English in the first assignment*”. However, some lecturers were disappointed because the test did not match their expectations. Sue noted that

*there isn’t a very strong correlation between the score that they have got in IELTS and what they can actually achieve and the reason for that is because well you can have a beautiful written structure around that (IELTS task) but this isn’t actually what you’ll do*”

Likewise, Samy said that, “*there are references that students have to make, analysis, and actually this is not there*” (in IELTS). He described that, “*this is not a test that would test your writing ability to write at doctoral or masters level xxxxxx (long pause) the issue is not the test the issue is the wrong use of the test*”, because, as Kate pointed out, “*The demands of writing a dissertation are different from the demands of writing here*” (in IELTS). She took the argument further and said that, “*in doing a doctorate there is a quality of thinking and argumentation that you need to be able to do*”, while the IELTS test is “*a much easier task than them looking at an academic paper and evaluating it*”.

(b) Support for international students

As the lecturers mentioned above, some international students were below their expectations so they needed to offer support for them to help them cope with the university demands. For example, John stated that, “*there is more time available to supervise international students*”. He explained that the type of support he would offer could be about using language for specific purposes because, for instance, the language of “*social science research is very technical*” and its terms are “*not always very identical to the everyday language use of the word xxx (pause) reliability for example*”. He would also: “*give detailed feedback on texts*”, “*work with a piece of draft material*”, “*reword things and make suggestions about how it could be written more convincingly*”, and ask students to be critical because, “*There is so much emphasis in the UK in higher degree courses in critical engagement*”. In doing so, he seemed to be aware of the fact that he should not rely on the IELTS scores because: “*if students are able to write it at IELTS 6.5 level that doesn't necessarily mean that they have the flexibility to deal with that technical language*”.

Abraham who appeared to be in favour of raising the IELTS scores noted that the lower the IELTS score of candidates the more support is needed on the part of receiving institutions. He said that, “*universities who accept students with 6 on IELTS should provide more support*”. This support according to him may be in the form of “*a one-to-one session, sometimes we give on job support*”. The limit of this support is when: “*the mistakes in writing a lot, the grammar is incorrect, the typos are a lot, and not being enough care to the writing this would let me lose track of what is being said and with the academic substance*” and, “*I will say no we need to give this for somebody to proofread*”.

Similarly, Kate would leave minimal linguistic errors for a proof-reader but she can still look at students' texts globally in aspects such as "*organization, in their development of logical arguments, their criticality and their sort of conventions in writing*". However, she added that, "*I don't think I can help them in English. I talk about the structure and the idea but I don't really talk about English*". Sue agreed that lecturers: "*don't proofread students assignments but most lecturers are very tolerant when it comes to that so they don't expect perfection*". Unlike the IELTS test where more attention is paid to language and structure, university lecturers according to Sue, "*are looking for the whole global thing not just on the writing but we are looking at writing in relation to the task that they are trying to achieve*".

Actually, Samy had a more meaningful philosophy about support and he thought that it was somebody else's job to do proofreading. He said that he did not like to "*correct these mistakes*" because "*I am not a language teacher*". Alternatively, what he would like his students to do in research was to follow the three words rule, namely "*depth, breadth and intellectual curiosity that's my personal philosophy*". He explained that,

*"Depth means when you study your area you study it so well that you will become an expert in it. Breadth means you connect between your area and other subjects. You make the links between the micro and the macro level. That's how you become successful. Intellectual curiosity all the seminars that happen all the talk the reading the context then you become a scholar. That's what a doctor student or master's student should be like"*.

Similarly, David had a different view of support. He viewed support from a cognitive as well as a social side. He emphasised that what international students should "*appreciate more is the importance of independent thinking*" rather than just "*repeating knowledge gained from textbooks*". He added that we want postgraduate "*students to synthesise*

*knowledge gained through scholarship with their own independent thinking*". Most importantly, he pointed out that universities should "*encourage us to find time to make it more of a social and cultural encounter rather than just a carefully timed and monitored academic relationship*". That is because he viewed teaching and supervising postgraduate students "*as a shared journey, but the atmosphere is too pressured and it is becoming too contractual in its nature*".

#### **5.4 Scoring validity question:**

- To what extent are the rating scales and criteria used consistently by markers?

To answer this question, the study used an open-ended questionnaire. The purpose was to find out about the degree of consistency when scoring the AWM in IELTS. The open-ended questionnaire aimed at investigating scoring validity of AWM in IELTS from the position of IELTS marker. The following section describes and reports the findings of this questionnaire.

##### **5.4.1 Markers' Questionnaire**

This questionnaire consisted of two parts: personal information and questions relating to the scoring of AWM in the IELTS test. Specifically, the first part contained 8 categories asking about the marker's age, gender, level of education, major, first language, nationality, relevant experience and training, and years of employment. The second part examined rating scales facets, such as: the frequency of referring to the scoring guide during the scoring process; the importance of training sessions; the reason for awarding candidates differently in different centres or occasions; the difference between scoring Task 1 and Task 2 in terms of text features; the features of the text that markers looked

for in both tasks while marking; their opinion on the validity and clarity of the scoring method; the frequency of asking for double marking; and the way they arrived to the final score.

It should be noted that results of second part were only made available to the researcher. This was due to the confidentiality of the IELTS test and the difficulty of obtaining IELTS markers to participate in the study. Therefore, the only examiner who was involved in the study did not want to be known or identified. Results of the second part are reported below.

Scoring validity in this study is concerned with the degree of consistency when scoring AWM in IELTS. This is because inconsistency in marking could affect the reliability of test scores (Shaw and Weir, 2007). When the IELTS marker in this study was asked about the frequency of referring to the scoring guide during the scoring process, her reply was that, *“I know them already so there is no need to have them with me while scoring.”* This may show that she used to have the rating criteria with her at the beginning of her career but as she was progressing through the job and perhaps taking training courses she internalised them and found that she no longer needed them. The IELTS marker in this study stressed the importance of training sessions by saying that they *“are essential for gaining consistency among markers”* because *“different markers may interpret the rating scales differently”*.

Based on this premise, candidates should not be awarded differently in different centres or occasions, *“because marking criteria are similar and the process of marking has been standardised”*. She added that *“Yet, it could be that some markers may tend to be lenient*

*while scoring*". This indicates that there are some attitudes, e.g., leniency or severity that may be challenging to overcome for markers despite the fact that they had training sessions.

In fact, double marking of candidates papers may help (Shaw and Weir, 2007) but it is not always called for because it could be time-consuming and expensive. Thus, single marking, as the marker in this study pointed out, could be adequate: "*Probably yes because IELTS is marked through an analytic scale*". By that she may have meant that IELTS markers can be more consistent when using analytic marking scales where criteria are predefined and each criterion has a separate score rather than just one global score in holistic marking. For instance, she gave a general yet useful description about the process of marking, especially how to arrive to the overall score in an analytic scale for marking. She described that, "*When I am marking the whole text, I give a band score for each criterion. I then add them up together and decide on the corresponding overall band score which would best describe the performance of the candidate*". She mentioned that there was a difference between scoring Task 1 and Task 2 in terms of text features where: "*Task 2 is longer and it carries more weight in terms of marks towards the overall score*". More specifically, she added that the features of the text that she looked for in both tasks while marking were: "*Task 1 coherence and cohesion, length, grammar and vocabulary. Task 2 the same criteria*".

More importantly, she warned that markers need to be aware of factors that may affect the consistency of scoring that are according to what she said: "*Lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring*".

Actually, these factors, as mentioned above, may not only be a threat to scoring validity but to the reliability of test scores. In addition, if the marking is not reliable there may be a risk of wasting the time and effort that was put to create a valid instrument (Alderson et al., 1995).

## **5.5 Conclusion**

This chapter has reported the results obtained from the participants in Phase I and Phase II of the study. It was divided into three parts according to the research questions of this study. The findings of the first question were about the students' awareness and perceptions of the IELTS test and the way they responded to the test tasks. The participants' perspectives of the degree of correspondence between IELTS tasks and university tasks were revealed in the findings of the second question. The findings of the last question showed the consistency of markers during the scoring process. The key findings of the study for each question will be presented in the following chapter according to the Phases of the study.

# CHAPTER 6

## **Chapter VI**

### **Discussion**

#### **6.1 Introduction**

This chapter aims to explore the links between the phenomena investigated in the study and the main themes that arose during the analysis. Results that were obtained from different data collection methods will be reviewed and summarised here in order to provide answers for the research questions. The established links will be discussed in the light of the themes with reference to other studies in the literature. The purpose is to show whether there are consistencies or differences between the results of the study and similar studies in the literature. The discussion of the findings will be presented thematically as follows: the stakeholders' perceptions about the test, the students' engagement with the test, and the connection between these perceptions and engagement.

#### **6.2 The key findings of the study**

This section reports the key findings of the study in two different phases with the participants, namely: the students, the marker and the lecturers. The first section reports the key findings of Phase one of the study in which students were given Pre-writing Test Questionnaires, Post-writing Test Questionnaires, Observations and Interviews. It also presents the findings of the Marker's Questionnaire. The key findings of Phase two are reported in the second section that consists of the students' Questionnaires and Interviews as well as the Lecturers Interviews. Key findings here are reported according to the

research questions (see Appendix I for a report of the key findings according to instruments).

## **6.2.1 Phase I**

### **6.2.1.1 Key findings for response validity question**

*a- What sense do students make of the tasks in the writing test?*

*b- And to what extent does this affect their performance?*

Generating valid responses from the candidates in test tasks is an important issue in measuring their language proficiency. The key finding for this question is that the students' interpretations of test prompts were seen to affect their performance in producing appropriate responses to the test tasks. The process of making meaning of test prompts required a careful analysis of the question in order to understand the social function of the task (i.e. genre) and then determine how the text should be composed. The students occasionally referred to the prompt of the tasks to make sure that they covered every aspect in the question. They were scanning for the keywords in the questions and as soon as they identified them and understood the requirements of the tasks they started writing. Some students indicated that the more detailed the question of the task the more ideas they will have. It seems that the delay time while responding to test tasks may have affected the use of certain test taking strategies. Moreover, due to time restrictions, some students were seen to be planning and revising as they were writing and very often their choice of test taking strategies was deliberate. That is to say, they had a metacognitive awareness of their writing strategies but they were not able to implement them under test conditions.

The students have approached the tasks differently and this could be why they spent different times while responding to test tasks. The first task was seen to present more challenge to most of these participants due to either time limitations, stress, difficulty in understanding the graph and the table, too much information in the graph and the table versus limited information in the rubric, or missing information. Likewise, the second task was also demanding to the rest of the students in terms of bearing more marks and words, doing two things simultaneously, and bringing outside information. Some candidates have shown that certain textual features in the prompts have aided them to meet two important criteria in marking academic writing tasks in IELTS, namely, Task Achievement (in Task 1) and Task Response (in Task 2).

#### **6.2.1.2 Key findings for scoring validity question**

*- To what extent are the rating scales used consistently by markers?*

In order for a test to have high scoring validity, markers need to be consistent in their application of the scoring criteria (Weigle, 2002). The key finding for this question is that the perceptions of the marker in this study showed that different markers may interpret the rating scales differently. As a result, some candidates may not be allocated correctly to their corresponding band score or level of performance because of leniency or severity of the marker. Training courses were perceived to be essential for gaining consistency among markers. Yet, the issue of leniency or severity might be challenging to overcome for markers despite the fact that they had training sessions. Therefore, the marker in this study warned that markers need to be aware of the following threats to scoring validity and to the reliability of test scores, namely: lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring.

## **6.2.2 Phase II**

### **6.2.2.1 Key findings for content validity question**

*- In what ways do academic writing test tasks correspond to the types of tasks students do at university?*

Content relevance and content coverage are two essential aspects for a test to claim content validity (Bachman, 1990). The key findings of this question is that the students reported that there was a mismatch between IELTS tasks and university tasks in terms of level of difficulty, variety of topics, styles, discourse mode, length, and language.

Similarly, the university lecturers in this study indicated that there was low level of correspondence between IELTS tasks and university tasks. The students indicated that university tasks were more academic, longer, and related to one's area of study, whereas the IELTS tasks were perceived to be more general, much easier, very structured, and shorter in terms of time and word count. For these reasons, they believed that the IELTS test lacked some essential skills needed for academic writing and therefore were doubtful of its representativeness.

Likewise, the university lecturers in this study expressed their concerns over the issue of readiness as university courses require high thinking and critical writing skills while IELTS would equip students with the technical elements. They indicated that IELTS alone is insufficient to prepare students for university. Therefore, most of them were against raising the IELTS scores because the test gives a rough measure of what a candidate is able to do.

### **6.3 The stakeholders' perceptions about the test**

This section discusses the perceptions of the stakeholders (students, lecturers and marker) about the test. It has four sub-themes under it. They are as follows: the students' perceptions of the value of the test, the stakeholders' perceptions of test awareness, the stakeholders' doubts about IELTS correspondence to university tasks, and the stakeholders' doubts about IELTS as a measure.

#### **6.3.1 The students' perceptions of the value of the test**

In this study, students considered the value of the IELTS test to be as a means of access more than as a measure of proficiency. Specifically, before they applied to any university abroad they decided on which country they would go to study. That is because this would determine which test to take according to its preference in the country where they were going to study. Almost all students were applying to UK or Australian universities and chose to study in countries where IELTS was the preferred test for admission except for one candidate who applied to US universities. It could be that these students knew about this preference of IELTS in those universities through the application forms where IELTS came first, or sometimes the only test accepted and the equivalent scores of other test were aligned to its scores. Arguably, the IELTS test is now gaining high recognition in the US as it is accepted in more than 3000 institutions as a proof of English language skills (IELTS, 2012a). Therefore, students who took the test and satisfied the minimum entry requirements had increased their chances of being accepted in more than one place.

Clearly, this indicates that the value of the IELTS test was seen in its preference by target universities rather than its high reliability or validity in measuring academic

proficiency. In other words, the majority of students seemed to be more likely driven by an institutional choice rather than a personal one regardless of what was their purpose for taking the test. This is consistent with the results of Smith and Haslett's (2007) investigation as they found that IELTS had a symbolic value beyond its rationale as a proficiency test because of its wide use and acceptance in high-stakes situations. This also confirms what Coleman et al. (2003) found in their study where both staff and students pointed out that the primary purpose for IELTS was to gain entry to university and its role as an indicator of language proficiency was a secondary consideration. In essence, the students in this study seemed to be affected by the social contexts which made an impact on their choice of IELTS test.

### **6.3.2 The stakeholders' perceptions of their test awareness**

Students in this study showed that they had a fairly high level of awareness of test content despite some few inaccuracies about the duration of the test. This finding was drawn from data collected before taking the mock test through the pre-writing test questionnaire. For example, they identified the difference between the first and the second writing task in issues such as length, type, difficulty, time, word limit, topic, style, questions, and relevance. They were also aware that their responses should be accurate, interesting, free from spelling mistakes, clear, well-organised, formal, coherent, focused, completed, correct, appropriate, and not necessarily related to answers in both tasks.

Gaining this knowledge about the test on the part of the students was perhaps seen to be essential as it may impact positively on their performance during the test. In other words, such high level of awareness could possibly boost their performance to successfully accomplish the writing task. This might be also seen in applying certain test-taking

strategies, such as time management. For instance, if they knew in advance the duration of the writing test and allocated time for each task, they would highly likely set priorities and have a predefined plan for each task completions, within the given time. This indicates that test preparation appears to be an important factor in increasing the candidates' knowledge of the test and helping them to master the essential skills needed to get the targeted IELTS score.

This confirms the findings of Mickan and Motteram, (2008) which showed that IELTS preparation courses tended to engage students into a process of socialisation into test-taking strategies and priorities in the test. They added that instruction targeted at raising awareness of test-taking behaviours, interpretation of task's prompts, as well as time management. Students in this study confirmed those findings as they preferred IELTS preparation courses (focus on the test only) to EAP courses (focus on developing academic English skills needed at university) in order to improve their scores. This is also consistent with Brown's (1998) findings who suggested that IELTS preparation courses provided students with more knowledge needed for score gain than the general EAP programme. This could be because students in the IELTS preparation courses would perhaps have specific internalised set of expectations about the requirements of the task which would help them accomplish the task.

However, IELTS preparation courses would provide less knowledge than EAP courses about academic writing needed at university (Green, 2007). The impact of IELTS preparation courses on score gain was seen to be more dramatic in writing than reading (Rao et al., 2003). Furthermore, IELTS preparation courses helped students to get ready for the test. Green (2007) confirmed this as he concluded that students who were enrolled

in IELTS preparation courses felt more confident about taking an IELTS test at the end of their courses than students in the EAP courses. Above all, Weir (2005) stressed that students should be fully aware before the test of the marking criteria as this will influence their planning and monitoring strategies. He gave an example where if the students felt that organisation was not as important as grammatical accuracy and therefore planning would be considered less important to them.

Similarly, the marker who participated in this study showed a high level of awareness of test content and scoring criteria. This was expected as IELTS markers are selected carefully according to the following criteria obtained from IELTS website (IELTS, 2012f):

- applicants must be native speaker (or a non-native speaker of an overall IELTS Band 9 – with a 9 in both Speaking and Writing modules);
- applicants must have a relevant TEFL qualification eg CELTS/DELTA;
- applicants must have at least three years relevant teaching experience; and
- applicants must have an undergraduate degree (or equivalent).

After becoming an IELTS examiner, further training is required every two years for re-certification in which examiners are assessed face-to-face by an accredited IELTS trainer (IELTS, 2012f). Training courses, as the marker in this study noted, were essential for gaining consistency among markers. That is because there could be some variations among markers in their performance while scoring the candidates' texts. It has been acknowledged that such variation in rater's performance can be kept within acceptable limits through rater training - a method which is widely used for this purpose (O'Sullivan

and Rignall, 2007). They concluded in their study that in order to yield high consistency and reliability among examiners, feedback in rater training has to be given repeatedly over a period of time rather than 'one shot' feedback. Therefore, the main concern of certification, as the marker in this study indicated, is to make sure that markers are rating to the standard because marking criteria are similar and the process of marking has been standardised. This means that there should be no variations in the awarding process as candidates should receive the same score wherever they are.

However, the marker mentioned that even after being trained and certified it could be that different markers may interpret the rating scales differently and also some markers may tend to be lenient while scoring. This variation and bias could mean that examiners may not reveal their actual performance during training. That is to say, they may seem consistent during training but as soon as training finishes this is no longer the case. It has been found in several studies that examiners become less severe during training (Furneaux and Rignall, 2007; Ruth and Murphy, 1988, and Weigle, 1998 ). This perhaps may have led McNamara (1996) to argue that "the traditional aim of rater training - to eliminate as far as possible differences between raters - is unachievable and possibly undesirable" (p.232). He suggested that self-consistency should be the alternative aim for any training programme for raters.

Yet, I would argue that in order for the test to be fair for every candidate regardless of where and when they take the test, there should be an inter-rater consistency, otherwise some candidates may not be allocated correctly to their corresponding band score or level of performance. Davies et al. (1999, p.200) emphasised that "[i]f candidates' chances of success on the test are liable to increase or diminish according to when or where they are

tested or who happens to be assessing them, then the test is unfair”. Fulcher and Davidson (2007) put it rightly in their discussion of standardisation as they argued that “consistency in language testing is about avoiding any variation in procedure that could threaten score meaning” (p.127).

Accordingly, if different IELTS markers would vary dramatically in giving scores for the same candidate the test will be unfair. Some students in this study mentioned that they tended to get high scores in IELTS in the receiving countries rather than their home countries. This is consistent with the results of Mickan and Motteram (2009) who found that some students felt that taking the IELTS test locally was much more challenging than taking it overseas. They added that there was a general impression that international centres tended to be more generous with the students in terms of getting high scores. If this were to be true then IELTS test results are unfair and unreliable. Hence, it could be that for this reason the marker in this study warned that “*[m]arkers need to be aware of several factors that may not be a threat to scoring validity but to the reliability of test scores, namely: lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring*”. Training programmes for markers as mentioned above claim to tackle these issues in order to improve the inter-rater reliability (similarity of the marks given by different examiners) and inter-rater reliability (the consistency in the severity of an examiner’s marking) among IELTS examiners (Shaw and Weir, 2007).

Despite the training that she received, the IELTS marker in this study gave different names for some of the criteria for scoring both tasks. For instance, she said that the features of the text that she looked for in both tasks while marking were “*Task 1 coherence and cohesion, length, grammar and vocabulary. Task 2 the same criteria*”. In

fact, the criteria are not exactly the same because Task 2 has task response instead of task achievement (IELTS, 2010a). In addition, the names for some criteria, though not so much important, were not accurate, such as Length (Task Achievement), Grammar (Grammatical Range and Accuracy), and Vocabulary (Lexical Resource) (IELTS, 2010a). Arguably, the difficulty in recalling the exact scoring criteria could be attributed to several factors:

- 1- The fact that she is retired and therefore no longer remembers those criteria exactly.
- 2- The fact that she was marking other Cambridge tests, as she told me, which had slightly similar criteria to those in IELTS.
- 3- It could be that examiners may tend to focus on certain criteria and neglect others which made her confused about the exact names.
- 4- It might be that markers feel confident about the scoring method and criteria, as the marker stated that, "*I know them already so there is no need to have them with me while scoring*". She believed that expert markers do not need to have the scoring guide during the marking process because they have already internalised it.

The findings of Furneaux and Rignall (2007) (on IELTS writing module) are consistent with the third factor. They found that raters may sometimes neglect some of the criteria while marking candidates' papers because some of the prescribed criteria were more difficult than others to apply. For example, they added that raters tended to refer to criteria such as Vocabulary and Sentence Structure or Task Fulfilment more than the criterion of coherence and cohesion. In this case, the reliability of test scores is at risk.

That is because the marker may not have given a full account for all of the different areas of performance of that candidate. Likewise, Cotton and Wilson (2011) found that examiner's marking of coherence and cohesion was less reliable than any other marking criteria of the second writing task in IELTS. Their findings indicated that the reason for this low reliability in marking was because the marking of coherence and cohesion was more difficult than any other criteria and that some examiners were ignorant of what these terms mean.

So, contrary to what the marker indicated with respect to the scoring guide, markers (especially those who are marking more than one type of test) may need to at least have a look at the marking criteria before embarking on the scoring process of any different test. Bachman and Palmer (2010, p.73) suggested that it is a good practice to ask raters to "review the rating criteria and benchmark examples before each rating session". That is because Wolfe et al. (1998) found that the most reliable and consistent markers were those who stayed closer to the scoring guide.

### **6.3.3 Stakeholders' doubts about IELTS correspondence to university tasks**

In this study, the students reported that there was a mismatch between IELTS tasks and university tasks in terms of level of difficulty, variety of topics, styles, discourse mode, length, and language. More specifically, they indicated that university tasks were more academic in terms of vocabulary, longer in terms of time and word count, and related to one's area of study in terms of topics. Moreover, they added that university writing demanded more effort because doing research would involve reading references and conducting surveys to collect data for the study. For these reasons, the students found

IELTS to be more general with respect to topics and vocabulary, much easier, very structured, and shorter in terms of time and word count. Nonetheless, the students noted that both tasks were similar in several areas, such as language, i.e. academic English, discourse mode, e.g. descriptive and argumentative, style, e.g. objective writing, structure, cognitive demands, e.g. analyse and write, organisation, grammar and vocabulary. In this regard, they indicated that IELTS writing test was somehow suitable to prepare students for university study but may not be sufficient on its own if students wanted to develop those skills (e.g. referencing, critical thinking, managing long papers or assignments, etc.) needed for academic writing.

Similarly, the university lecturers who participated in this study indicated that there was low level of correspondence between IELTS tasks and university tasks. That is because university courses require high thinking and critical writing skills, while IELTS would equip students with the technical elements (e.g coherence and cohesion, etc.). Moreover, the demands of writing an assignment at university are different from those in the IELTS writing test. Therefore, they suggested that IELTS alone is not enough to prepare students for university and thus universities should offer more support during the academic programmes. Meanwhile, they added that IELTS is best suited for preparing undergraduate students for year one because this is the kind of activities they will be doing. According to what these lecturers said, the test could be misused with the current practice as it is testing postgraduate students who are applying to either a Master or a PhD course.

The findings of this study are consistent with the findings of Moore and Morton's (2007) study which investigated the authenticity of Task 2 in the IELTS academic writing

module and examined its relevance to university writing. They found that the genre was somehow similar in both academic and general domains. Yet there were some differences, such as:

- 1- The use of prior knowledge as the basis for writing in the IELTS tasks, compared with the prescription of a variety of research processes in the university assignments.
- 2- A restricted range of rhetorical functions in the IELTS items (with a focus on hortation), compared with a diversity of functions in the university tasks.
- 3- An emphasis on ‘real world’ entities (situations, actions, practices) as the objects of enquiry of IELTS items compared with a greater focus on abstract entities (theories, ideas, methods) in the university tasks. (Moore and Morton, 2007, p.197).

The study also found that the IELTS academic writing tasks were opinionated in terms of style which made them more related to specific public non-academic genres. This has led the researchers to suggest that these types of tasks are not a suitable model for university writing. Likewise, Banerjee (2000) contended that IELTS academic writing modules are relatively short compared to university assignments. Hence, they fall short of providing stakeholders with the necessary information about the ability of the candidates to produce coherent, appropriate and accurate texts which are similar in terms of length to those produced at university.

Additionally, the results of this study are in accordance with the students’ concerns in Green’s (2007) study with respect to Task one in AWM in IELTS and its relationship to

university tasks. He pointed out that his students found that the first task required them to make a description only whereas university tasks would require deep interpretation and conclusions from given information. Moreover, the task was seen to be irrelevant to some disciplines in terms of graphical information. Furthermore, the task items (e.g. graph, table, chart, etc) were sometimes inconsistent as they yielded either too much or too little information for 150 words task. In sum, the students have found that IELTS tasks shared some common features of target language tasks. Nonetheless, they were doubtful of the representativeness of IELTS tasks as they may not exactly mirror the content and skills needed to undertake university tasks.

Perhaps, it is worth reminding ourselves of Weigle's (2002) argument about timed writing tests. She argued that a timed writing test with a topic unknown to students in advance is "relatively inauthentic in at least four ways" (p.52). These are as follows:

- 1- In academic writing contexts (e.g. at university), students have the advantage of knowing the topic and dealing with it through reading, speaking and listening before writing and therefore have background knowledge and schemata within which to write.
- 2- Most academic writing is not timed or speeded and students can make use of outside resources, and revise and edit their work before submission.
- 3- The audience for writing tests is frequently unknown to the candidates while in most academic writing contexts the audience is the instructor and that gives

them an idea of how their texts should look like in order to meet the expectations of the marker.

- 4- The primary judgment in scoring criteria for most academic writing is on accuracy of content, rather than on the appropriateness of the organisation or the use of language (Weigle, 2002).

However, Taylor (2007) contended that,

“IELTS Writing tasks are not designed primarily to simulate the sort of university-level writing tasks which test takers will encounter in their studies. Instead, tasks are designed to be accessible to a wide range of test takers and to reflect features of writing activities that are already familiar to candidate from their previous study experience as well as some features of writing they may encounter in their subsequent study” (p.482-3).

Apparently, Taylor believes that authentic IELTS writing tasks are those tasks which mirror both the candidate background knowledge and the knowledge needed in the target situation. Yet, Davies (2008) seems to disagree with Taylor as he has a different view of authenticity in IELTS tasks. He asserted that, “tests simply cannot be authentic: what they can (all they can) do is to simulate authenticity” (p.69). This simulation of authenticity implies that full authenticity is impossible but the test should reflect as many features of the target situation as possible (not the candidate background knowledge). He pointed out that this is the position that the IELTS test adopted ultimately during the mid 1990s in the second revision. Based on this premise, the IELTS tasks do not claim authenticity in the strong sense of the definition but a simulation of those tasks that test takers would be expected to encounter or perform in the TLU domain outside the testing contexts. This understanding of authenticity is in line with the theoretical developments in language

teaching and learning in what has been known as communicative language testing in which the focus of testing has shifted from testing behaviours to testing abilities (Davies, 2008). That is to say, the aim of communicative language tests such as IELTS is to test language use rather than language knowledge.

#### **6.3.4 Stakeholders' doubts about IELTS as a measure**

The previous section above discussed the participants' views about the level of correspondence between IELTS tasks and university tasks. This section discusses an important and strong claim made about the test itself as a measure of academic proficiency. Arguably, this claim is about the issue of readiness. Taylor (2007) stressed that the IELTS test was “designed to test *readiness to enter* the world of university-level study in the English language and the ability to cope with the demands of that context immediately after entry” (p.482, her italics). However, some lecturers, in this study, expressed their concerns over the issue of readiness that is claimed by the IELTS test as they expected students to be at a level in English where they can survive at university but they were disappointed because the test did not match their expectations.

As a result, they were doubtful that students would be able to cope with the demands of the university. This finding is in line with what Davidson (2005) noted about teachers' impressions about the IELTS test. He pointed out that “[t]he success of the IELTS at determining students' readiness for baccalaureate study is contentious,” simply because “many teachers are not convinced that the academic format of the exam is sufficiently rigorous to act as a sufficient predictor of students' success at university” (ibid, p.51). According to the lecturers in this study this could be because the students would come to

university with a preparation about a general approach to academic writing which would hardly enable them to produce the sort of complex academic writing language needed for doing assignments at a university level. In other words, producing a short piece of writing during the test would not test the candidate's ability to write an academic case in a university course.

For this reason, these lecturers complained that the IELTS test gives a rough measure of what a candidate is able to do and therefore it may be difficult to generalise on the students' performance in these tasks. That is to say, in claiming readiness to enter university IELTS may not be clear about how the students are going to cope with university demands upon gaining entry. In a call to clear this confusion, Tony Green (2009, personal communication) proposed that "[p]erhaps we need a clearer statement from IELTS about what is meant by 'readiness' ". The findings of this study are also in line with Rea-Dickins et al.'s (2007) study who found that "[t]he assertion of readiness for postgraduate implicit in the IELTS score (which afforded entry to programmes) emerged as problematic in both students' and tutors' accounts" (p.118). Furthermore, the findings of this study are consistent with O'loughlin's (2008) study who found that the staff (both academic and administrative) in his study had a "scepticism about the validity, reliability and "trustworthiness" of IELTS scores and an unrealistic expectation about their power to predict academic success" (p.145).

Likewise, the students in this study indicated that they were lacking essential academic (research) skills needed for success at university and that the IELTS test offered them limited support in this regard. Despite the fact that they have achieved the minimum requirements in IELTS scores in order to satisfy the entry requirements in the receiving

institutions the students felt that they were not yet ready to undertake university tasks. Presumably, they seem to challenge the claim about readiness that IELTS academic writing test assesses “whether a candidate is ready to study or train in the medium of English at an undergraduate or postgraduate level” (IELTS Handbook, 2007, p.2). Therefore, the students as well as the lecturers suggested that more support (e.g. developing academic writing, research skills, critical thinking, etc) is needed in order to cope with the university demands. This means that universities should not expect the IELTS test solely to equip students with such skills needed for academic success. That is because the use of IELTS as an “indicator of subsequent academic performance is limited” (Hyatt and Brooks, 2009, p.39). Generally, Douglas (2010) added that “language tests alone cannot tell us whether someone will be a good postgraduate student” but rather tests only show “that there is a likelihood that the person possesses a level of language ability that will be, or not be, adequate for subsequent tasks” (p.34).

In fact, when universities found that the students could not cope (or were not ready) with the demands of the university despite achieving the minimum entry requirements, they raised the minimum IELTS scores for entry in order to make sure that they have better students. Receiving institutions have the right to determine the cut-off score which meets their expectations and that is why there is a variation in the minimum requirements for entry into university in terms of the IELTS scores. Nonetheless, it was recommended that receiving institutions should take into account other important individual factors when making decisions about admission.

According to the IELTS Handbook (2007, p.5) receiving institutions “should also consider a candidate’s IELTS results in the context of a number of factors, including age

and motivation, educational and cultural background, first language and language learning history”. In addition, the disclaimer in the application form for the IELTS test confirms that “IELTS is not designed to be the sole method of determining admission or employment for the test taker” (IELTS, 2011, p.4). Yet, I would argue that if IELTS partners wanted test users to be aware of the limited uses of the test, they should have put this statement somewhere else (because the disclaimer is rarely read) where it can be clearly seen by test users with a font size significantly bigger than the current 7.5. This is because “there is evidence to suggest that in some institutions, proficiency test scores are used in somewhat cruder fashion as . . . ‘pass-fail indicators’” (Brindley and Ross, 2001, p.150). Accordingly, it is incumbent for the receiving institutions to “consider how language ability, individual factors and academic requirements fit together to ensure more dependable admission decisions” (Chalhoub-Deville and Turner, 2000, p.538).

In the UK context, it was found that universities set their cut-off scores at IELTS 6 or 6.5 (Hawkey, 2006). Yet, these cut-off scores might have changed since then due to the current recommendations to increase the minimum scores. More specifically, Davies (2008) suggested that

[a]n institution which wishes to encourage foreign students may choose to establish a low cut-off and take more students, thus risking a higher proportion of false positives, while an institution with a highly competitive entry and a restricted number of places may wish to set a high cut-off and ensure fewer misses but more false negatives (p.24).

This may indicate how much faith receiving institutions put in the overall scores and how they define them which may sometimes lead them to give less attention to other factors recommended by the IELTS Handbook above.

However, lecturers in this study were not happy about raising the IELTS scores because it wouldn't ensure that only good students would be accepted into university as the difference in performance would be small and may not be noticed. This is in line with the findings of Brown's (2008) study as he warned that

Increasing the minimum language requirement to IELTS 7 may not guarantee improved language ability, and besides, such a change in recruitment policy in a volatile and competitive market would only happen in an ideal world, we don't live in an ideal world (p.90).

Additionally, Feast (2002) acknowledged that raising the IELTS scores may "result in unacceptably high losses of international students for very modest GPA gains" (p.84). Moreover, these findings are consistent with the results of Sawyer and Singh's study (2011) who reported an inconsistency problem in the IELTS scores as the participants in their study argued that if all 6.5 students were like their best 6.5s students, the IELTS would be a reliable measure, while this is not the case and thus "raising the standard changes nothing" (p.117).

Interestingly, one of the lecturers mentioned that the reason why they were going to raise the IELTS scores was because they wanted to compete with other universities in a certain top group in terms of standards as they wanted to be within the acceptable minimum scores in that group in order to survive the competition. This confirms the findings of McDowall and Merrylees (1998) who noted that there was some sort of consultations among institutions when setting the cut-off scores. They stated that, "there could be a strong temptation for each institution to set the policy according to what it perceives the other equivalent universities/colleges have done" (p.134). In addition, findings of this study are in line with what Chalhoub-Deville and Turner (2000) noted about the practice

of setting cut-off scores. They indicated that “[o]ften, institutions do not carry out any systematic local investigations to decide upon cut-off scores, but instead base their decisions on hearsay and what other institutions are doing” (ibid, p.532).

Most importantly, Shohamy (2001) gave another example about the use of cut-off scores to serve administrative purposes. She warned that cut-off scores should not be lowered or raised because of the number of places available for students at university but instead they should be based on the level of language needed for success. Spolsky (1998) noted that the boundaries between scores are artificial:

the sad truth [is] that most users act as though the score a student receives is a real absolute measure with an immediate non-contextualised interpretation, and as though there is a real difference between a student who scores 597 and one who scores 601 (p.12).

Similarly, Rea-Dickins et al. (2011) found that IELTS test scores were seen by receiving institutions as a “true” indicator of language proficiency and as “hard” and accountable evidence (p.262). However, they argued that “the students gaining entry to the programme through the hard criterion of IELTS may not be ready for or suited to the programme in terms of other criteria” (ibid, p.279).

Briefly, a high score in IELTS does not necessarily mean that the candidate is ready for a particular programme of study because some lecturers noticed that there was low correlation between IELTS scores and what a candidate can achieve. That could be because the interpretation of readiness from the perspective of the receiving institutions may not often match the abilities which the test claims to measure. In a study that investigated the stakeholders’ definition of the word ‘success’ for determining the

appropriate IELTS band score for admission, Golder et al. (2009) found that stakeholders defined the word ‘success’ as taking part in courses, working in groups for assignments, and interacting with other students and instructors. However, in this case, they argued that the IELTS test, which measures language tasks was not consistent with this broad definition of success. In other words, the IELTS definition of readiness did not meet the stakeholders’ expectations of readiness for university tasks.

Furthermore, the marker in this study revealed some marking habits that some markers may sometimes apply which might have a negative impact on the consistency of the scoring process. For example, she said that different markers may interpret the rating scales differently. It may be argued that the reason for such variation in performance among markers and even sometimes for the same marker is because of the complexity of scoring written performance (Mickan, 2003). He described that raters need to semiotically interpret the criteria for marking, make sense of the written texts and assigning marking criteria to text features in the texts. More specifically, Brown (2006) reported that the examiners in the study found an “overlap between the scales, a lack of clear distinction between levels, and the inference-based nature of some criteria” (p.41). This could be the reason why variation exists in the scoring of candidates’ performance. Similarly, study by Lumley (2002), in which he investigated the processes by which 4 trained raters made scoring decisions when using an analytic scale to score 48 written texts by ESL learners found that it was difficult for raters to reconcile their impression of the text, the specific features of the text, and the wordings of the rating scale. He suggested that raters can be successful in yielding consistent scores if they receive

adequate training and assistance in the form of guidelines to deal with problems arising during the scoring process.

In addition, Sakyi (2000) investigated the thinking processes and criteria used by raters in evaluating written texts. His analysis revealed the different reading styles adopted by raters as they read the texts: “(1) focus on errors in the text, (2) focus on essay topic and presentation of ideas, (3) focus on the rater’s personal reaction to text, and (4) focus on the scoring guide” (p.129). This suggests that not all raters focus on the scoring guide as there were other aspects which diverted their attention from it. The marker in this study pointed out that leniency or severity of certain markers could be another factor affecting the consistency of the scoring process among markers despite the fact that they had training sessions.

The marker in this study reported that single marking could be adequate in terms of validity and reliability of scores because IELTS is marked through an analytic scale. Indeed, it was noted that an analytic scale would allow markers to attend to the specific features of the texts, avoid impressionistic decisions leading to rater bias and dismiss norm-referencing (Shaw and Falvey 2008). However, Fulcher (2010) argued that this may give rise to a halo effect because “when raters are asked to make multiple judgements they really make one and this affects all other judgements”. The different interpretation practices of candidates’ performance and scoring criteria on the part of the raters may together affect the reliability and the validity of test scores which as a result may risk having candidates who are not allocated appropriately to their corresponding band score or level of performance. Therefore, Uysal (2010) contended that “despite the reported high reliability measures, in such a high-stakes international test, single marking

is not adequate” (p.2). Similarly, Davies (2007) asserted that “Cambridge ESOL have made serious attempts to develop procedures that will assure stakeholders that IELTS Speaking and Writing are reliable measures, but it does seem that the doubts will continue as long as single marking is retained” (p.84).

#### **6.4 The students’ engagement with the IELTS writing test**

This section discusses the students’ engagement with the test. There are four sub-themes under this section. These include: the students’ perceptions of test difficulty and duration, the students’ perceptions of test prompts and instructions, the perceived test taking (Students) and scoring strategies (Markers), and expertise in second language (L2) writing.

##### **6.4.1 The students’ perceptions of test difficulty and duration**

In this study, students took IELTS writing mock tests (i.e. Task 1 & Task 2). Most of them indicated that the first task presented a challenge to them because there was too much information in the rubric which caused confusion, and it was difficult to establish a connection between the table and the chart. Yet, some of them found the words of the questions of both tasks to be free of any problems causing misunderstanding or confusion. Put differently, the confusion was not in understanding the wording of the question, though it was a bit demanding, but in handling too much information in the table and the graph. As a result, the students’ performance was not at its best due to spending too much time on understanding the requirements of the task.

This outcome is consistent with the findings of Brossell (1986) who found that students composed higher quality essays when they were given prompts with medium level of

information load rather than those with full specification with unnecessary repeated information (high level) or no specification (low level). Kroll and Reid (1994, p.233) listed three types of prompts, namely: a bare prompt (states the entire task in direct and simple terms for the candidate), a framed prompt (presents a situation which acts as a frame for the interpretation of the task) and a text based prompt (presents a text which requires students to write an essay or use in their own writing). The first task in the IELTS test can be considered as having a text based prompt, whereas the second task can be described as having a framed prompt. Also, the findings of this study are in accordance with the findings of O'Loughlin and Wigglesworth (2007) who examined task difficulty (IELTS Academic Writing Task One) in terms of amount and presentation of information to candidates. The researchers found that tasks with less information were more complex than those with more information. It should be noted that the researchers meant by more information here adequate or moderate (at a medium level) information rather than too much information. Recently, (Yu et al., 2011, p.35) suggested that it is important to “consider the information density, or data points of the graphs, as information density clearly has impact on how and what test takers can extract from the graphs”.

Accordingly, it seems that the more unnecessary information student are given in a prompt the more confusion there would be, and the less time would be available to students in order to write. Clearly, the prompts of the first task in the mock IELTS test in this study had a negative impact on the students' performance in terms of the cognitive load at the beginning of the task. More specifically, students were asked to summarise, select, report and then make comparisons. Seemingly, this may look easy for students as

they only had to transfer information from the table and the graph into their actual report. Yet, it would require a lot of thinking on the part of the students about which particular information should be included or excluded.

Because the students noted that there was too much information in the first task they suggested that the solution was either to reduce the amount of information given in the task or allow them to have more time to do the task. The average time that they asked for was 25 minutes to complete the task which was an increase of 5 minutes to the current task. That is because the nature of the task required them to read (the information in the table and the graph) and then write. For this reason, it was suggested that, “if a writing task is to be based on reading or the description of a graph, for example, more time will be required” (Weigle, 2002, p.102). This suggested timing of the task may be used by students to organise their information in the form of drafts. Yet, most of them reported that they were unable to make drafts, especially in the first task, and some of them could not even complete the task because of time restrictions.

The second task, however, provided the students with sufficient time to complete the task. Nonetheless, it was demanding to some students in terms of doing two things at a time, i.e. discussing both sides of the arguments and then giving their opinions about a certain issue, and bringing outside information to support their ideas by evidence and examples. These two factors may have added extra cognitive demands and pressure on the test takers, especially if we knew that the second task carried more weight in marking than the first task and required more words. Apparently, depending on background knowledge when responding to a particular prompt to construct an argument would require students to “generate both the ideas and the language and is thus more

challenging than a prompt in which the content is provided by stimulus material”

(Weigle, 2002, p.94).

So, the second task involved a lot of thinking as it was not just knowledge telling (e.g. first task) but also knowledge transforming activity (Bereiter and Scardamalia, 1987).

Despite this different level of difficulty, the students were generally happy with the current order of the tasks and considered the first task to be a warm up for the second one.

Most importantly, the above discussion indicated that little variations (low, medium or high information) in the prompts of the tasks can either affect positively or negatively the cognitive and linguistic abilities of the students. This is in line with what Mickan et al. (2000) noted in their study about of response validity of the IELTS writing tasks. They found that, “[b]oth the readability of prompts and the subjects’ skills in the formulation of appropriate responses impact on the response validity of the test items” (p.44).

In sum, the first task was seen to present more challenge to most of these participants due to either time limitations, stress, difficulty in understanding the graph and the table, too much information in the graph and the table versus little information in the rubric, or missing information. On the other hand, the second task was also demanding to the rest of the students in terms of bearing more marks and words, doing two things at a time, and bringing outside information.

#### **6.4.2 The students’ perceptions of test prompts and instructions**

In this study, students valued the importance of instructions. For example, they reported that they work as a facilitator and make the task easier to understand, and they can help those students who will be taking the test for the first time without prior preparation. Yet,

the students emphasised that they needed more guidance in order to focus their attention on the important aspects of the task that in turn may help them to allocate the appropriate time for each part, an issue that was always of concern to them. It seems that these students felt that the prompt of the task was lacking focus and clarity. That is because it was perhaps a bit too general. Obviously, the role of instructions is to “specify the task the test taker is expected to complete, and the test-taker performance is enhanced when there is a greater appreciation of the task or of what is required of the test taker” (Shaw and Weir, 2007, p.65).

However, the instructions in the first task that the students took in the mock exam did not, according to them, meet these conditions. More specifically, the students were asked to do four things: summarise, select, report and then make comparisons. The challenge was not in doing more than one thing at a time but in understanding the requirements of the task. For example, Abdu pointed out that “*say summarise ok summarise but tell me a specific thing to do*”. Moreover, the students had a difficulty in establishing the relationship between the graph and the table. This may mean that the instructions were not sufficiently detailed for these students so that they were clear about what they were expected to do, as students are not assessed on their ability to understand test instructions but rather on their responses to assessment tasks (Bachman and Palmer, 2010).

In essence, what the students needed from the instructions was to help them structure the task in the way that would match the markers’ expectations. That could be why they occasionally referred to the prompt of the tasks to make sure that they covered every aspect in the question. Also, they needed instructions to facilitate the task and not to make it more complicated so that they spend adequate time in order to understand the

requirements of the task. In other words, they wanted clear task prompts in order to encourage appropriate cognitive processing needed to enhance their performance in the test. That is because students have different interpretations of task requirements as they respond to the same task differently (Ruth and Murphy, 1984). With this in mind, it seems that the less guidance students get from the task prompt the more room there might be for misinterpretation of task requirements.

#### **6.4.3 The perceived test taking (Students) and scoring strategies (Markers)**

In a timed test such as IELTS, students need to consciously make use of certain test taking strategies in order to assist them to accomplish the intended target goal/s easily within the given time. Before we start talking about the strategies that the students used in this study it is important to know why the timed test is fit for purpose for these students. Impromptu (or timed) tests can test a large number of students (that is why it is called a large scale test) in a short time where test scores are used to discriminate among test takers performance, it is the most appropriate test for students from different programmes and experiences. The assessment will likely have a positive rather than a negative washback, and it focuses on the individual while ensuring that the test was taken without outside assistance (Weigle, 2002). In fact, there are four criteria on which candidates are assessed on the IELTS academic writing test. These include: Task Achievement (for task 1), Task Response (for task 2), Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy (IELTS, 2010a).

In this study, students' responses indicated that they would focus on these criteria differently at the beginning of the writing test. That is to say, students had different

priorities before taking the test for these criteria. Nevertheless, they were keen to first meet those criteria that would seem to be less controversial to examiners. For example, most students indicated that they would pay attention to task achievement while writing. It seems that this criterion was the easiest among other criteria as it was only to do with word count. So, writing 150 words (with relevant, accurate and appropriate information) in the first task meant that they have secured a band score or so in that task. Hamp-Lyons (1991a, p.52) pointed out that “[e]ach writer needs both guidance as to what is important about this writing task and what qualities will be valued”. Clearly, the students seized the chance in getting the most out of the task achievement criterion.

Other criteria in the test were achieved by using a number of strategies such as planning, revision and brainstorming. Indeed, the majority of the students reported that they would consider planning and revision strategies to be the most important strategies in the writing test. Also, they stressed the importance of brainstorming in enhancing the quality of their writing. Interestingly, what the students noted with regards to these strategies was somehow in line with what Bachman and Palmer (1996) considered to be the essential characteristics of an interactive task.

According to them, an interactive task has the advantage of engaging students in metacognitive strategies, which are part of their strategic competence, that help them to undertake a range of processes including **goal setting** (involves identifying the language use tasks or test tasks, and choosing one or more tasks from a set of possible tasks and deciding whether or not to complete the task); **assessment** (involves assessing the characteristics of the language use task to determine the desirability and feasibility of accomplishing it, assessing the elements of topical and language knowledge to determine

whether the necessary elements are available, and monitoring and evaluating the correctness and appropriateness of response - utterance or interpretation - in accomplishing the chosen tasks); and **planning** (involves the formulation of one or more plans for implementation as a response to the task). Apparently, it seems that Bachman and Palmer (1996) used goal setting to refer to brainstorming, and assessment to mean revision. This indicates that the students have made effective use of their strategic competence for better management of their cognitive and linguistic resources to accomplish the task. This is also indicative of the interactiveness of IELTS academic writing tasks as they engaged the students in this study in almost all the metacognitive processes characterising interactive tasks.

Most importantly, the aforementioned findings of my study showed that students were conscious of choosing the writing strategies. That is to say, the cognitive processing activities addressing and linking various stages during the production of writing did not to a large extent happen by chance. These findings are in line with what Shaw and Weir (2007) noted about scoring criteria and writing processes. They pointed out that when the scoring criteria are made known to candidates before the test they will have an impact on the processing of the task in terms of planning, monitoring and revision.

On the other hand, the students in this study mentioned that time limitation was an issue of concern to them when applying some of these writing strategies. For example, some of them described that they could not brainstorm in the first task because of time restrictions. Other students were able to brainstorm in both tasks and found that the first task did not need as much brainstorming as the second one. There was also a variation in the students' responses about the frequency of going back to correct grammar mistakes,

style and organisation during and after the writing process. This could be attributed to the kind of the task, its requirements, and the time limitations associated with it. These findings support the conclusion that Weigle (2002) reached when reviewing research on writing. She explained that “individuals use different cognitive, rhetorical and linguistic strategies when they are faced with tasks that vary according to topic, purpose, and audience” (p.69). This may also suggest that where one strategy would seem to be helpful to somebody, it could be, however, less useful and time consuming for somebody else due to certain restrictions or requirements.

Likewise, unfamiliar topics made the process of generating ideas through brainstorming more difficult and had a negative impact on the students’ performance during the test. Truly, topical knowledge is so important to an extent that “[t]he content knowledge required for completing a particular task will affect the way it is dealt with” (Weir, 2005, p.75). In this study, for instance, Amro complained that the most problematic aspects to him were: “*The time and the topic because the topic will be new to me and may need longer time*”. Similarly, Fatin mentioned that she was “*Shocked by unusual topics, which I don’t have many words about and lose the linking between the ideas*”. It could be that when she prepared for the test she expected a certain type of topics to appear but she was disappointed as she was given different topics. These concerns about topical knowledge were raised in detail in the previous chapter with more examples from other students. Essentially, these findings are consistent with Read (1990) who argued that students would produce better essays when writing about a familiar topic than an unfamiliar one. In addition, these findings are similar to Xiaoyue’s (2010) study results, who found that

topic familiarity gave rise to a slightly more accurate performance, especially in lexical diversity and lexical sophistication.

Most importantly, these findings reveal that the topics of the IELTS writing tests may neither be interesting nor accessible to all test takers. In fact, this contradicts what the IELTS claims about the construct of the academic writing test. It claims that, “[t]he issues raised are of general interest to, suitable for and easily understood by candidates entering undergraduate or postgraduate studies or seeking professional registration” (IELTS Handbook, 2007, p.8). Therefore, in a test such as IELTS which claims that it is testing general academic proficiency in English, tasks that require specific background knowledge should be avoided in order to achieve the above claim, especially if we know that topics have an effect on test scores (Papajohn, 1999). In sum, lack of familiarity with the topic and time pressure may, as shown above, hinder the students from showing their best performance in the test in one way or another, and therefore we may risk having an insufficient evidence of what the students are able to accomplish in the target situation under such restrictions.

In the same vein, the marker in this study reported that different markers may have different scoring strategies. That is to say, they differ in their application, in terms of order, of the four marking criteria for writing mentioned already. This is in line with Eckes’s (2008) classification of raters. He stated that the difference among raters can be seen

(a) in the degree to which they comply with the scoring rubric, (b) in the way they interpret criteria employed in operational scoring sessions, (c) in the degree of severity or leniency exhibited when scoring examinee performance, (d) in the understanding

and use of rating scale categories, or (e) in the degree to which their ratings are consistent across examinees, scoring criteria, and performance tasks (p.156).

#### **6.4.4 Expertise in second language (L2) writing**

The above discussion highlighted the various strategies that students used during the IELTS writing test tasks and their perceptions about them. This section tries to find possible links between the components of writing strategies and the different levels of language proficiency for these second language writers. That is to say, there are certain cognitive processes that characterise performance at different levels and which help us to distinguish between expert and novice writers. In this study, students have frequently mentioned that planning and revision were the most important strategies they used repeatedly in the test. Such description of performance may tell us something about their level of expertise. According to Bereiter and Scardamalia (1987) they are perhaps skilled writers as they allowed more time for planning before starting to write, generating ideas and revising, whereas less skilled writers stop and translate more and revise words, rather than larger discourse sections during their writing due to insufficient time given to planning.

This indicates that expert writers have different composing processes than novice writers. In this study, this difference was seen in terms of effective use of writing strategies that the students made use of during the mock writing test. This difference may also be seen in the high level of complexity and lexical density of skilled writers' texts as a result of planning (Xiaoyue, 2010). Perhaps, these strategies may have been transferred from the first language to the second language context (Silva, 1993). Interestingly, it was

noted that “unskilled L2 writers are unlikely to engage in organisation” due to the high cognitive load that they go through while attempting to encode their “thoughts in linguistic form so that the resources available for building cohesion and coherence may be severely limited” (Shaw and Weir, 2007, p.38).

## **6.5 The connection between perceptions and engagement**

This section tracks the change of perspectives on the part of the students who were followed in two phases in this case study. Below is a detailed description of this process.

### **6.5.1 Change of perspectives**

In this study, the data was collected from three kinds of participants, i.e. students, marker and lecturers. The focus of this section is on the students. Their progress was tracked in two phases over a period of time in different settings. The aim of this section is to shed light on the change of the students’ perspectives as they progressed over a period of time. Not surprisingly, they even sometimes had different perspectives in the same phase as different data collection methods had revealed during the analysis. For instance, students in the pre-writing questionnaires in the first phase showed that they had high level of awareness of test content. However, in the post-writing questionnaires in the same phase, which was given right after the mock test, they reported that they were less familiar with the content of the writing tasks, especially of the first one. My interpretation of this is perhaps due to the unusual representation of data in the first task, as the students indicated, where there was a graph and a table. It could be that when the students prepared for the test they had only tasks with either a graph or table and not both. In addition, it seems that familiarity with the content meant different things to these

students. Some have associated familiarity with the test format or response to the question by following a specific genre. On the other hand, other students have seen familiarity as being able to know the content of the task by checking the given information in the task against their prior knowledge about the subject or the content of the task.

Likewise, students' perspectives changed with regards to the instructions in the first task. In the post-writing test questionnaire, they valued the importance of instructions and they found the words of the questions of both tasks to be free of any problems causing misunderstanding or confusion. However, they complained in the interview that the current guidelines, especially in the first task, were not helpful as they left a number of things implicit such as the relationship between the graph and the table. My interpretation of this is possibly that they referred to the instructions only in the post-writing questionnaire whereas in the interview they meant the instructions with relation to the test item or question in the first task. The reason for this could be because the students were exposed to a high cognitive overload during the test while they were trying to figure out the relationship between the graph and the table. It seems that the students were trying to match the information that they knew about the test either from preparation materials or test handbook with that in the test rubric. Apparently, they were surprised with this huge amount of information about the test task as they were used to see less than that in those preparation booklets. So, instructions were of vital importance to these students, but some specific information regarding the relationship between the graph and the table, was missing. Instead, this was left for the students to figure out and this consequently caused high cognitive overload to them.

Furthermore, the students in this study had different perspectives on different occasions about the duration of the second task. In the post-writing test questionnaire, almost all of the students reported that the allotted time for the second task was sufficient. Yet, they pointed out in the interview that time limitation was one of the obstacles they faced in the second task. The reason why they had inconsistent views about the time allotment in the second task might be because they felt that they could not make full use of brainstorming. That is to say, it seems that they were generally happy with the current time allocation but when it comes to making effective use of writing strategies, e.g. brainstorming, the second task would need more time for that. Seemingly, what the students said above suggests that there might be a tension here between the testing under timed conditions and what they normally do in untimed writing contexts.

Similarly, students changed their perspectives as they progressed towards the second phase of the study with respect to the issue of representativeness of IELTS tasks. In the first phase, they said in the interview that the test focused on language ability needed for survival in the target language use context. However, in the second phase, they mentioned in the questionnaire that IELTS writing test tasks may not be sufficient on their own if students wanted to develop those skills (e.g. referencing, critical thinking, managing long papers or assignments, etc.) needed for academic writing. My interpretation of this is that at the first phase they took the test and they thought it would prepare them for university. Yet, in the second phase, when they started their university courses, they discovered that the test did not match their expectations in this regard.

## **6.6 Summary of the chapter**

This chapter has synthesised the findings of the two phases of the study. It then discussed the views of the different stakeholders (students, lecturers and marker) with reference to the literature. It presented their perspectives on the construct validity of the test in three types of validity, i.e. content validity, response validity, and scoring validity. This has given us an insight of what the test is testing, the way students are making sense of test tasks, the level of correspondence between the test tasks and the university tasks, and the consistency of markers in using the rating scales. It is these factors that affect the students' performance in the test or the validity of their scores either positively or negatively. It should be noted that these three types of validity are referring to one concept that is construct validity but they have been separated here for descriptive purposes only. Briefly, this chapter has discussed the following major themes in the study. These include:

### **1- The stakeholders' perceptions of the test**

This theme discussed, firstly, the students' perceptions of the value of the test and how they were affected by the social contexts that made an impact on their choice of IELTS test. Secondly, it talked about the stakeholders' perceptions of their test awareness where the students as well as the marker showed high level of awareness of test content and scoring criteria despite some few inaccuracies about the test. This was followed by a discussion of the stakeholders' doubts about IELTS correspondence to university tasks where the students and the lecturers reported that there was a mismatch and a low level of correspondence between IELTS tasks and university tasks in terms of level of difficulty,

variety of topics, styles, discourse mode, length, and language. Finally, this theme took into consideration the stakeholders' doubts about IELTS as a measure because some lecturers, in this study, expressed their concerns over the issue of readiness that is claimed by the IELTS test and they were doubtful that the students would be able to cope with the demands of the university. Lecturers were also not happy about raising the IELTS scores. Likewise, the students in this study indicated that they were lacking essential academic (research) skills needed for success at university and that the IELTS test offered them limited support in this regard. Furthermore, the marker in this study revealed some marking habits that some markers may sometimes exhibit which might have a negative impact on the consistency of the scoring process.

## **2- The students' engagement with the IELTS writing test**

This theme started with a discussion of the students' perceptions of test difficulty and duration where most students indicated that the first task presented a challenge to them because there was too much information in the rubric which caused confusion, and it was difficult to establish a connection between the table and the chart. It then talked about the students' perceptions of test prompts and instructions where students found the instructions and the prompts of the test to be an important factor contributing to their understanding of task requirements. This was followed by a discussion of the perceived test taking (Students) and scoring strategies (Markers) where students have showed that they consciously made use of certain test taking strategies in order to assist them to accomplish the intended target goal/s easily within the given time. Also, the marker revealed that different markers have different marking strategies and they may interpret the scales differently. Finally, this theme tackled expertise in second language (L2)

writing which showed that there are certain cognitive processes that characterise performance at different levels and which help us to distinguish between expert and novice writers.

### **3- The connection between perceptions and engagement**

This section tracked the change of perspectives on the part of the students who were followed in two phases in this case study and gave a detailed description of this process.

The next chapter will conclude the discussion by providing a summary of the whole dissertation. It will also state the limitations of the study and will suggest some implications and recommendations for future research.

# CHAPTER 7

## **Chapter VII**

### **Conclusion:**

### **Summary, Significance, Limitations, Implications, and Recommendations**

#### **7.1 Introduction**

The previous chapters have illustrated a number of issues related to the design of the study including a presentation of its rationale and significance, a description of its context, a review of literature in the relevant fields, a description of the methods and methodology, and an analysis and a discussion of the findings. This chapter will focus on transforming the stakeholders' perceptions into statements that describe their importance in informing the current assessment practice of the academic writing module in the IELTS test. That is to say, it attempts to demonstrate the contribution of the study to the literature around language testing in general and performance assessment of academic proficiency in second language writing through IELTS in particular. This chapter will also outline the limitations of the study and will propose some implications and recommendations for future research.

#### **7.2 Summary of the study**

After reviewing the relevant literature on IELTS, the current study noted that previous studies have been fragmented in terms of addressing the issue of validity of the IELTS AWM (i.e. focusing on one type of validity). Therefore, this study argued that the validity

of the IELTS AWM may be affected by one or more of these three factors, namely the test, the candidates and the markers. For this purpose, this study has investigated the construct validity of the IELTS test from the perspective of the stakeholders' in order to inform us why some Saudi candidates fail to cope with the demands of university study despite the fact that they have achieved the minimum entry requirements in the IELTS test. With respect to the content validity of the test, this study found that the tasks in the academic writing test to a large extent do not reflect the kind of tasks candidates are likely to encounter at university. Moreover, this study pointed out that response validity, on the part of students who may not have understood the rubric of the tasks, is another important factor affecting the students' performance. Furthermore, this study suggested that scoring validity has a significant effect on the students' scores because of the inconsistency of markers during the scoring process as they may have sometimes failed to assign the students to their corresponding level of proficiency. This study has emphasised that it considers validity as a unitary concept where these types of validity constitute construct validity (Messick, 1989) and they have been treated separately here for investigation purposes only. The following sections will give a brief summary of the likely causes for the students' failure at university.

### **7.2.1 The stakeholders' perceptions about the mismatch between IELTS tasks and university tasks**

In this study, the students indicated that there was a low level of correspondence between IELTS tasks and university tasks in terms of level of difficulty, variety of topics, styles, discourse mode, length, and language. Specifically, they reported that university tasks were more academic in terms of vocabulary, longer in terms of time and word count, and

related to one's area of study in terms of content. In addition, they added that university writing demanded more effort because doing research would involve reading references and conducting surveys to collect data for the study.

Because of these reasons, the students found IELTS to be more general with respect to topics and vocabulary, much easier, very structured, and shorter in terms of time and word count. Nevertheless, the students noted that both tasks were similar in several areas, such as language, i.e. academic English; discourse mode, e.g. descriptive and argumentative; style, e.g. objective writing; structure; cognitive demands, e.g. analyse and write; organisation; grammar and vocabulary. In this respect, they indicated that IELTS writing test was largely suitable to prepare students for university study but may not be sufficient on its own if students wanted to develop those more advanced skills (e.g. referencing, critical thinking, managing long papers or assignments, etc.) needed for academic writing.

Likewise, the university lecturers who participated in this study pointed out that there was a mismatch between IELTS tasks and university tasks. That is because university courses require high thinking and critical writing skills while IELTS would equip students with the technical elements (e.g. coherence and cohesion, etc.). Additionally, the demands of writing an assignment at university are different from those in the IELTS writing test. Hence, they suggested that IELTS alone is not enough to prepare students for university and therefore universities should offer more support during the academic programmes. They added that IELTS is currently best suited for preparing undergraduate students for year one because this is the kind of activities they will be doing. According to what these

lecturers said, the test could be misused with the current practice as it is testing postgraduate students who are applying to either a Master or a PhD course.

To sum up, the students have found that IELTS tasks shared some common features of target language tasks. Nonetheless, they were doubtful of the representativeness of IELTS tasks as they may not exactly reflect the content and skills needed to undertake university tasks.

### **7.2.2 Stakeholders' doubts about IELTS as a measure**

The above section summarised the participants' views about the level of correspondence between IELTS tasks and university tasks. This section summarises the stakeholders' perceptions about the issue of readiness that is claimed by the IELTS test, i.e. being able to cope with the demands of the university level courses after successfully meeting the minimum entry requirements in IELTS. Yet, some lecturers, in this study, were doubtful that students would be able to cope with the demands of the university. That is because they expected the students that they taught to be at a level in English where they can survive at university, but unfortunately they were disappointed because the test did not match their expectations. In this case, it seems that the test to these lecturers cannot be considered a sufficient predictor of success at university.

The lecturers, in this study, explained that this could be because the students would come to university with a preparation about a general approach to academic writing that would hardly enable them to produce the sort of complex academic writing language needed for doing assignments at a university level. Put differently, asking students to produce a short piece of writing during the test would not test the candidate's ability to write an academic

case in a university course. For this reason, these lecturers complained that the IELTS test gives a rough measure of what a candidate is able to do and therefore it may be difficult to generalise on the students' performance in these tasks. In other words, by claiming readiness to enter university IELTS may not be clear about how the students are going to cope with university demands upon gaining entry.

In the same vein, the students in this study reported that they were lacking essential academic (research) skills needed for success at university and that the IELTS test offered them limited support in this regard. Although they have achieved the minimum requirements in IELTS scores in order to satisfy the entry requirements in the receiving institutions, the students felt that they were not yet ready to undertake university tasks. Based on this finding, the students as well as the lecturers proposed that more support is needed in terms of developing academic writing, research skills, critical thinking, etc., in order to cope with the university demands. This means that universities should not expect the IELTS test solely to equip students with such skills needed for academic success.

Currently, when universities found that the students could not cope (or were not ready) with the demands of the university despite achieving the minimum entry requirements, they raised the minimum IELTS scores for entry in order to make sure that they have better students. However, the lecturers in this study were not happy about raising the IELTS scores because it wouldn't ensure that only good students would be accepted into university, as the difference in performance would be so small and may not be noticed. Interestingly, one of the lecturers mentioned that the reason why they were going to raise the IELTS scores was because they wanted to compete with other universities in a certain

top group in terms of standards as they wanted to be within the acceptable minimum scores in that group in order to survive the competition.

In sum, a high score in IELTS does not necessarily mean that the candidate is ready for a particular programme of study because some lecturers noticed that there was low correlation between IELTS scores and a candidate's performance at university. That could be because the interpretation of readiness from the perspective of the receiving institutions may not often match the abilities that the test claims to measure.

Furthermore, in expressing her doubts about IELTS as a measure, the marker in this study indicated that different markers may interpret the rating scales differently and this may have a negative impact on the consistency of the scoring process. As a result, the inconsistency of scoring may cause sometimes a variation among markers in the marks given to the same candidates. Moreover, the marker in this study pointed out that leniency or severity of certain markers could be another factor affecting the consistency of the scoring process among markers despite the fact that they had training sessions. Furthermore, the marker reported that single marking could be adequate in terms of validity and reliability of scores because IELTS is marked through an analytic scale. Briefly, the different interpretation practices of candidates' performance and scoring criteria on the part of the raters may together affect the reliability and the validity of test scores which as a result may risk having candidates who are not allocated appropriately to their corresponding band score or level of performance.

### **7.2.3 Tracking the change of perspectives**

This section tracked the change of perspectives on the part of the students who were followed in two phases in this case study. In this tracking, the researcher tried to make a connection between the students' perceptions of the test and their engagement with it. Below is a detailed description of this process.

#### **7.2.3.1 Change of perspectives**

This section focused on the students whose progress was tracked in two phases over a period of time in different settings. The aim of this section is to shed light on the change of the students' perspectives as they progressed over a period of time. Interestingly, sometimes some students had different perspectives even in the same phase as different data collection methods had revealed during the analysis. For example, the students showed a variation of responses in the first phase in terms of familiarity with the test content between the pre-writing questionnaires and the post-writing questionnaires. Similarly, the students' perspectives changed in the post-writing test questionnaire and the interview with regards to the instructions in the first task.

Furthermore, the students in this study had different perspectives in the post-writing test questionnaire and the interview about the duration of the second task. Apparently, their perceptions suggest that there might be a tension here between the testing under timed conditions and what they normally do in untimed writing contexts. Likewise, the students' perspectives in the first phase changed as they progressed towards the second phase of the study with respect to the issue of representativeness of IELTS tasks.

### 7.3 Significance of the study

The present study has some contributions to those who are involved in second language writing performance assessment in Saudi Arabia. These contributions include:

#### 7.3.1 *In terms of originality:*

- 1- To the best knowledge of the researcher, there is no study that has investigated the Saudi students' perspectives about the construct validity (i.e. content validity, response validity and scoring validity) of the Academic Writing Module (AWM) tasks in the IELTS test from a socio-cognitive perspective.
- 2- In addition, there is also no case study as such that investigated the stakeholders' perspectives (i.e. candidates, lecturers and markers) about the construct validity of the AWM in the IELTS test from an interpretivist point of view using triangulation and followed the same students on two occasions.

#### 7.3.2 *Significance for stakeholders:*

- 1- This study informs the *stakeholders* of the limitations of the IELTS test as a sole preparatory tool to university study and raises their awareness that it is not designed to assess the full range of academic language genres and skills needed for university study.
- 2- This study provides an insight for *test takers* regarding the test taking practices and strategies.
- 3- This study points out that having a high level of awareness of test content through preparation may help *test takers* to focus on the important aspects that are valued

by test examiners and may also help them to allocate the appropriate time for each part of the task.

- 4- This study assists *teachers* in language centres who are preparing students for the IELTS test in knowing which strategies should be promoted most often and which ones should be used less often.
- 5- This study draws the attention of the *receiving institutions* to the risk of raising the IELTS scores each time, as it would not guarantee that only good students would be accepted at university. Instead, universities should offer their support in terms of research skills, academic language, etc. instead of raising the sub-score or the overall score which may not make a big difference in performance.

#### *7.3.3 Significance for test providers:*

- 1- This study provides an insight for *test providers* into the way test takers make sense of test tasks.
- 2- This study draws the attention of *test developers* to including model examples of writing texts in the IELTS preparation materials written by the IELTS examiners so that the expected response could be highly predictable.

#### *7.3.4 Significance for IELTS examiners:*

- 1- This study provides an insight for *IELTS examiners* into the factors such as: the lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring that may cause a threat to scoring validity.

- 3- This study stresses the importance of training sessions for *IELTS examiners* in order to have a high level of consistency in scoring among them and to have a full understanding of the scoring criteria.

#### **7.4 Limitations of the study**

Though every effort has been made to improve the quality of this research, several limitations have to be acknowledged. These limitations are related to both the design and methodology of the study.

This study was exploratory in nature and collected data through multiple methods, i.e. questionnaires, interviews and observations. Arguably, the possible use of verbal protocols with the students where they could verbalise their thoughts or think aloud as they were writing during the mock test would add more insights about their cognitive processing and how they made sense of the test tasks. The researcher did not wish to use this method simply because he wanted to simulate the real test in almost every aspect in which the interruption of the students is prohibited. Instead, the researcher decided to use observations, post test questionnaires and interviews right after the test to compensate for this loss and get rich data from the students.

Another limitation concerns the sample of the study. Because this study was multi-site and conducted on two occasions, it was difficult to maintain the same number of students in both phases. That is to say, only 8 students out of 15 wished to continue in the second phase of the study. The researcher was aware of the likelihood of this attrition and that was why he included 15 students in the first phase of the study. Yet, this drop on the part

of the students did not have a great effect on the results of this case study. Similarly, access to IELTS markers was difficult because of the confidentiality of the exam. Therefore, only one retired marker participated in the study by filling out a questionnaire. Furthermore, it was not possible to conduct other data collection methods such as the interview or the observation because the marker did not wish to be known or identified. Nonetheless, it is noteworthy that the perceptions of this IELTS examiner allowed the researcher to learn about some issues regarding the scoring of the AWM in IELTS. Without these perceptions, it would have been difficult for the researcher to understand how the marking was done. The next section will outline the implications of this study.

## **7.5 Implications of the study**

This study has a number of implications that may inform those who are involved in second language writing performance assessment. It is crucial to note that the following implications and recommendations need to be understood in the light of the limitations of the study mentioned earlier and in the context where it was conducted. The following sections give a detailed explanation of the study's implications to stakeholders, test providers as well as the methodological implications of the study.

### **7.5.1 Stakeholders**

This study has several implications to stakeholders or test users. For test takers, there are some implications regarding the test taking practices and strategies. In this study, students reported that planning and revision strategies were the most important strategies in the writing test. Although they had limited time available during the test, they made use of them by time management and were revising and planning as they were writing. This

indicates that students who want to sit for the writing test should pay more attention to these strategies. Having a high level of awareness of test content through preparation may help students to focus on the important aspects that are valued by test examiners and may also help them to allocate the appropriate time for each part of the task.

This knowledge may also help teachers who are preparing students for the IELTS test to know which strategies should be promoted most often and which ones should be used less often. Most importantly, the effective use of each strategy should be prioritised over the type of strategy. In addition, test takers should start with the second task in the writing test because it has more marks and according to the students in this study it was less demanding in terms of cognitive processing.

For receiving institutions, the implications of this study is that raising the IELTS scores each time would not guarantee that only good students would be accepted at university. Despite the fact that some universities are competing with each others in terms of high standards, i.e. wanting to be in a certain position or a group, universities should nonetheless understand that there is a certain level where most students can start from, e.g. 6 or 6.5. At this score or level, universities should offer their support in terms of research skills, academic language, etc., instead of raising the sub-score or the overall score which may not make a big difference in subsequent performance.

In sum, developing a greater understanding of the IELTS construct, i.e. its content and assessment criteria could also assist test users to make efficient use of the test and set realistic expectations when making inferences of its scores because, “the ultimate responsibility for appropriate test use and interpretation lies predominantly with the test

user” (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999, p.112). However, stakeholders should bear in mind that the IELTS test alone is not sufficient to prepare students for a university study as they need to be oriented to the other writing genres and to the academic language used in research contexts which is different from that in IELTS.

### **7.5.2 Test providers**

This study has also some implications for test providers. The perceptions of the stakeholders in this study showed that there was a mismatch between the IELTS tasks and the university tasks. On this basis, it was suggested that if writing test tasks were to reflect real life situations in measuring communicative ability, they should be designed by test developers to account for every possible response that would be accepted in target contexts and not just the one that existed in the mind of the examiner. Accordingly, test developers should emphasise, in the IELTS preparation booklets, their own desired model examples that would be highly appreciated by IELTS examiners and to give less attention to those examples written by candidates. In doing so, the test would become more transparent to them through test preparation materials and the expected response could be highly predictable. Uysal (2010, p.6) stressed that, “communicative aspects of writing rather than strict rhetorical conventions should be emphasized in the IELTS writing test”.

In addition, in this study, the students pointed out that some students may successfully get the required overall score but they may sometimes fail to meet the minimum requirements of one or more sub-scores. Therefore, it was suggested that in this case students should repeat just the modules that they failed in and not the whole test. The reason for such concern was because it was thought that test scores (either the overall

score or the sub-score) might fluctuate after repeating the test. Also, it is far less expensive for students to repeat one module instead of repeating the whole test. Furthermore, the students proposed that there should be specific modules for each discipline as they sometimes faced difficult questions that were not related to their field of study. The predecessor of IELTS, which was ELTS, used to have six modules representing different disciplines. However, it was criticised for being too general rather than too narrow (Davies, 2008).

For the IELTS examiners, the implication of this study is that they need to be aware of several factors that may not be a threat to scoring validity but to the reliability of test scores, namely: lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring. It was emphasised that the lack of expertise can be overcome by training, which would also help markers to be consistent with each other in terms of scoring. Yet, it was pointed out that the leniency and the severity of examiners could not be eliminated entirely by training but reduced to a tolerable level. Moreover, this study suggested that in order to avoid the problems of misinterpretation of scoring criteria, markers should have constant training as well as a focus on scoring one test. That is because an overlap between scoring criteria of different tests might exist and this may affect the quality of scoring.

Most importantly, IELTS test providers should tell test users what the half scores between each band mean. For example, in the current band scale for the IELTS test it is not specifically defined what a 5.5, 6.5, 7.5 mean. What is defined is the whole band score and the definition of the half band scores is left to test users to infer its meaning. Luoma (2001) argued that the main reason for construct definition is to enable test

developers to “say what the test scores mean” (p.103). This definition of half band scores is important because, “validity resides in the scores on a particular administration of a test rather than in the test *per se*” (Weir, 2005, p.12).

### **7.5.3 Methodological implications of the study**

This study is qualitative and thus the interest was in depth of coverage by collecting rich data about the participants in their natural settings. To do this, the researcher in this study had to let the participants speak for themselves in order to be able to give a full description of the phenomenon under investigation in that context and gain a new insight during the iterative exploration process. In this case study, data were collected and triangulated using three different methods including open-ended questionnaires, semi-structured interviews and semi-structured observations. Most importantly, the implication of this study is that triangulation here was not just in methods (e.g. questionnaires, interviews and observations), but in data (using multiple sources of information from different people, i.e. candidates, IELTS marker and university lecturers), and theory through the socio-cognitive framework (by triangulating four theories about writing, i.e. writing as a cognitive practice; writing as a social construction practice; writing as a social interaction activity; and writing as a sociocultural practice). These methods, participants, and theories complement each other in an attempt to add rigour to the data being explored.

Another implication of this study is the use of a socio-cognitive framework for writing to understand the phenomenon under investigation. The focus here is on the process as well as the product of writing. That is to say, in this study, the researcher was interested in capturing cognitive processes and social practices that may have an influence on the

performance of the students. Hence, the candidate lies at the heart of the assessment process rather than the instrument. According to Taylor (2006), this enhances the socio-cognitive approach to writing assessment “to be a more ‘person-oriented’ than ‘instrument-oriented’ view of the test- ing/assessment process than earlier models/frameworks” (cited in Shaw and Weir, 2007, p.242).

Furthermore, in this study, the students were followed in two phases and different contexts. The students’ perspectives have changed while describing the university context after the commencement of their programmes. For instance, in the first phase of the study in Saudi Arabia they had high expectations because they thought that the test would prepare them for university study with no need to take any further academic language courses for research. Yet, they were disappointed because the test offered them limited support in this regard. The implication here is that if accurate description of international students’ perceptions about their predictive performance is sought then it is a good practice to conduct a study on two occasions where the students are followed over a period of time.

## **7.6 Recommendations for future research**

The investigation of test validity is one significant aspect in language tests, especially in high stakes tests such as IELTS. Validation of tests is an on-going practice and has always been the interest of test developers, stakeholders, and researchers and it should still be so as long as we have language tests to offer. Construct validity lies at the heart of the validation process and the question about it, is possibly the most important one to consider at that stage (Davies et al., 1999).

Firstly, it is recommended that studies about the construct validity of the AWM in IELTS could be large scale and have bigger representative samples for the sake of drawing generalisations about the context. Those studies could take into consideration the variety of nationalities, backgrounds, contexts, levels of proficiency in English and ages of international students sitting for the IELTS test.

Another suggestion for future research is that construct validity has been regarded as a unitary concept consisting of more than one type of validity (Messick, 1989). Based on this modern view of validity, Weir (2005) listed five types of validity in a framework to conceptualise writing test performance: context validity (traditionally known as *content validity*), cognitive validity (this is very similar to the traditionally known *response validity*), scoring validity, consequential validity and criterion-related validity (see Appendix III). This study has focused on only three types of validity, i.e. content validity, response validity, and scoring validity. However, the investigation of the construct validity of the AWM in IELTS in the light of these five types of validity suggested in Wier's (ibid) socio-cognitive framework could allow future studies to have fuller understanding about the stakeholders' perceptions about the test.

A further area of interest is that the current study addressed the AWM in IELTS only. Future research, however, might be interested in including more than one skill or all of the four skills. Moreover, future research might wish to cover the two modules, i.e. general module and academic module of the IELTS test with a focus on one skill or all of the skills.

Most importantly, future research might be interested in investigating the impact of IELTS in terms of the financial issues on students repeating the test more than once. The price of taking the test has risen dramatically in the recent years. I have met some students who failed to meet the minimum entry requirements and they have repeated the test more than three times, and sometimes five times. Those students were complaining about the current test fees which are around £125-£130 depending on the selected test centre (IELTS Website, 2012a) and not to mention the preparation test booklets which are expensive too.

In the case of Saudi students, the Saudi Arabian Cultural Bureau compensates them for the cost of the test for two times only. If they were not successful, they have to pay for the cost themselves each time they want to sit for the test. Tellingly, the economic side of the IELTS test may have an impact on the lives of prospective students and thus future research may wish to examine its influence on them.

### **7.7 Final conclusion to the study**

In investigating the validity of the AWM of the IELTS test from the perspectives of the stakeholders, this study makes a contribution to the growing body of research about second language writing assessment in large scale/high stakes tests, especially the IELTS test, which is the focus of the study. The questions of this study have tackled the construct validity of the IELTS test, which is comprised of three critical aspects, i.e. content validity, response validity and scoring validity. The researcher employed three data collection methods, i.e. questionnaires, interviews and observations, which were then analysed qualitatively through NVivo. The findings of this study have given us an insight into how stakeholders, especially the Saudi students, perceived the construct validity of

the IELTS test. Their perceptions revealed many aspects about the test, chiefly among them, their doubtfulness regarding the issue of readiness that is claimed by the IELTS test, as well as its low correspondence to university study. Recommendations and implications for future research were made at the concluding chapter and it was advised that they should be seen against the limitations of the study.

# REFERENCES

- Adler, P., and Adler, P. (1987). *Membership Roles in Field Research*. London: Sage Publications.
- Adler, P., Adler, P. and Rochford, E. (1986). The Politics of Participation in Field Research. *Journal of Contemporary Ethnography*, 14(4), 363-376.
- Agar, M. (1991). The right brain strikes back. In N. Fielding and R. Lee (Eds.), *Using Computers in Qualitative Research* (pp. 181-194). Newbury Park: Sage Publications.
- Alderson, J. (1991). Bands and Scores. In J. Alderson and B. North (Eds.), *Language Testing in the 1990s: The Communicative Legacy* (pp. 71-94). London & Basingstoke: Macmillan Publishers.
- Alderson, J., and Banerjee, J. (2002). *State of the art review: Language testing and assessment (Part 2)*. *Language Teaching*, 35(2), 79-113.
- Alderson, J., Clapham, C., and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J., and Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14(2), 115-128.
- Al Fraidan, A. (2011). *Test-taking strategies of EFL learners on two vocabulary tests*. Lap Lambert Publications.
- Allen, P. (2009). *Definitions of Language Testing*. Retrieved October 18 2012, from <http://languagetesting.info/whatis/lt.html>.

Ambert, A-M., Adler, P. A., Adler, P., and Detzner, D. F. (1995). Understanding and Evaluating Qualitative Research. *Journal of Marriage and the Family*, 57(4), 879-894.

American Educational Research Association, American Psychological Association, and National Council of Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Anderson, J., Reder, L., and Simon, H. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5-11.

Applebee, A. (2000). Alternative Models of Writing Development. In R. Indrisano and J. Squire (Eds.), *Perspectives on Writing: Research, Theory, and Practice* (pp. 90-110). Newark, DE: International Reading Association.

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

Bachman, L. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671-704.

Bachman, L., and Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bachman, L., and Palmer, A. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.

Baker, C., and Johnson, G. (1998). Interview talk as professional practice. *Language and Education*, 12(4), 229-42.

- Banerjee, J. (2000). *Using English language screening tests in your institution*. A paper presented at the BALEAP Professional Issues Meeting, Nottingham Trent University.
- Barritt, L., Stock, P., and Clarke, F. (1986). Researching practice: evaluating assessment essays. *College Composition and Communication*, 37(2), 315–27.
- Beare, S. (2002). *Writing Strategies: Differences in L1 and L2 Writing*. Retrieved March 8 2010, from <http://www.llas.ac.uk/resources/paper/1292>.
- Beare, S., and Bourdages, J. (2007). Skilled writers' generating strategies in L1 and L2: An exploratory study. In G. Rijlaarsdam et al. (Ed.), *Studies in Writing Vol. 20, Writing and Cognition: Research and Applications* (pp. 151-161). Amsterdam: Elsevier.
- Berman, R. (1994). Learners' transfer of writing skills between languages. *TESL Canada Journal*, 12(1), 29-46.
- Berninger, V., Fuller, F. and Whitaker, D. (1996). A Process Model of Writing Development Across the Life Span. *Educational Psychology Review*, 8(3), 193-218.
- Börner, W. (1987). Writing in the foreign language classroom: thoughts on a model. In W. Lörcher, and R. Schulze, *Perspectives on language in performance, studies in linguistics Vol. 2, literary criticism and language teaching and learning* (pp. 1336–1349). Tübingen: G. Narr.
- Bosher, S. (1998). The composing processes of three South East Asian writers at the post-secondary level: An explanatory study. *Journal of Second Language Writing*, 7(2), 205–241.

- Braddock, R., Lloyd-Jones, R. and Schoer, L. (1963). *Research in Written Composition*. Urbana, IL: National Council of Teachers of English.
- Breen, M. (1985). Authenticity in the language classroom. *Applied Linguistics*, 6(1), 60–70.
- Brindley, G., and S. Ross. (2001). EAP assessment: issues, models and outcomes. In J. Flowerdew and M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.
- Britton, J., Burgess, T., Martin, N., McLeod, A., and Rosen, H. (1975). *The Development of Writing Abilities*. London: Macmillan Education Ltd.
- Britton, J., Martin, N., and Rosen, H. (1966). *Multiple Marking of English Compositions: An account of an experiment*. Schools Council Exams Bulletin, 12. London: HMSO.
- Brooks, E. (1985). *Case Studies of the Composing Process of Five “Unskilled” as a Second Language Writers* (Unpublished Doctoral Dissertation). New York University.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341–366.
- Brossell, G. (1986). Current Research and Unanswered Questions in Writing Assessment. In K.L. Greenberg, H. S. Weinder and R. A. Donovan (Eds.), *Writing Assessment: Issues and Strategies* (pp. 168-182). New York: Longman.
- Brown, A. (2006). An Examination of the Rating Process in the Revised IELTS Speaking Test. *IELTS Research Reports*, (Vol. 6, pp.41-70). IELTS Australia: Canberra and British Council: London.

- Brown, J. (1991). Do English and ESL faculties rate writing samples differently? *TESOL Quarterly*, 25(4), 587–603.
- Brown, J. (1998). An investigation into approaches to IELTS preparation, with particular focus on the academic writing module of the test. In S. Wood (Ed.), *IELTS Research Reports*, (Vol. 1, pp. 20-37). Sydney: Australia.
- Brown, L. (2008). Language and anxiety: An ethnographic study of international postgraduate students. *Evaluation and Research in Education*, 21(2), 75-95.
- Brualdi, A. (1999). *Traditional and Modern Concepts of Validity*. ERIC Clearinghouse on Assessment and Evaluation. Washington DC. Retrieved April 8 2010 from <http://www.ericdigests.org/2000-3/validity.htm>.
- Bryman, A. (1984). The Debate about Quantitative and Qualitative Research: A Question of Method or Epistemology? *The British Journal of Sociology*, 35(1), 75-92.
- Cambridge ESOL (2011). *IELTS Academic Writing Description*. Retrieved July 22 2011 from: <https://www.teachers.cambridgeesol.org/ts/exams/academicenglish/ielts/academicwriting>.
- Campbell, D. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297-312.
- Canale, M. (1983). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera (Ed.), *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application* (pp.107-123). Clevedon: Multilingual Matters.
- Canale, M., and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Journal of Applied Linguistics*. 1(1), 1-47.

- Cannell, C., and Kahn, R. (1968). Interviewing. In G. Lindzey and A. Aronson (Eds.), *The handbook of social psychology*, (Vol. 2, pp. 526-95). New York: Addison-Wesley.
- Carmines, E., and Richard, A. (1982). *Reliability and Validity Assessment*. Beverly Hills, CA: Sage.
- Carroll, J. (1975). *The teaching of French as a foreign language in eight countries*. New York: John Wiley and Sons.
- Centeno-Corte's, B., and Jime'nez Jime'nez, A. (2004). Problem-solving tasks in a foreign language: The importance of the L1 in private verbal thinking. *International Journal of Applied Linguistics*, 14(1), 7–35.
- Chalhoub-Deville, M., and Turner, C. (2000). What to look for in ESL admission tests: Cambridge Certificate exams, IELTS, and TOEFL. *System*, 28(4), 523-539.
- Charge, N., and Taylor, L. (1997). Recent Developments in IELTS. *ELT Journal*, 51(4), 374-380.
- Charmaz, K. (1983). The grounded theory method: An explication and interpretation. In R. Emerson (Ed.), *Contemporary field research* (pp.109-126). Boston: Little Brown.
- Charney, D. (1984). The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview. *Research in the Teaching of English*, 18(1), 65-81.
- Cohen, A. (1994). English for academic purposes in Brazil: The use of summary tasks. In C. Hill and K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174–204). New York: Longman.

- Cohen, A. (2012). Test Taking Strategies. In C. Coombe, P. Davidson, B. O'Sullivan, and S. Stoyhoff (Eds.), *The Cambridge guide to assessment* (pp. 96-104). Cambridge: Cambridge University Press.
- Cohen L., Manion, L., and Morrison, K. (2008). *Research Methods in Education*. (Sixth Edition). Abingdon: Routledge.
- Coleman, D., Starfield S., and Hagan A. (2003). The attitudes of IELTS stakeholders: student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions. *IELTS Research Reports* (Vol. 5, pp. 159-235). Canberra: IELTS Australia.
- Collins, P. (1991). Learning from the outsider within: The sociological significance of Black feminist thought. In M. Fonow and J. Cook (Eds.), *Beyond Methodology* (pp. 35-59). Bloomington: Indiana University Press.
- Cook, T., and Campbell, D. T. (1979). *Quasi- experimentation: Design and analysis for field settings*. Chicago: Rand McNally.
- Cotton, F., and Conrow, F. (1998). An Investigation of the Predictive Validity of IELTS amongst a Group of International Students studying at the University of Tasmania. In S. Wood (Ed.), *IELTS Research Reports*, (Vol. 1, pp. 72-115). Sydney: Australia.
- Cotton, F., and Wilson, K. (2011). An investigation of examiner rating of coherence and cohesion in the IELTS Academic Writing Task 2. *IELTS Research Reports*, (Vol. 12, Paper 6). IELTS Australia: Canberra and British Council: London.
- Cresswell, A. (2000). The role of portfolios in the assessment of student writing on an EAP course. In G. Blue (Ed.), *Assessing English for Academic Purposes* (pp. 205-220). Bern, Switzerland: Peter Lang European Academic Publishers.

- Creswell, J. (1994). *Research Design Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.
- Creswell, J. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. New Jersey: Merrill Prentice Hall.
- Cripper, C., and Davies, A. (1988). *ELTS Research Report 1 (i): ELTS Validation Project Report*. Cambridge: The British Council/UCLES.
- Crotty, M. (1998). *The Foundations of Social Research*. London: Sage.
- Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39(1), 81–141.
- Cumming, A. (1990). Metalinguistic and ideational thinking in second language composing. *Written Communication*, 7(4), 482–511.
- Cumming, A. (1996). *Building models of adult ESL/EFL writing instruction*. Paper presented at Second Language Acquisition and Writing Seminar, University of Southampton, July 12–13.
- Cumming, A. (2001). Learning to Write in a Second Language: Two Decades of Research. *International Journal of English Studies*, 1(2), 1-23.
- Cumming, A., and Riazi, A. (2000). Building models of adult second-language writing instruction. *Learning and Instruction*, 10(1), 55–71.
- Davidson, P. (2005). Using authentic assessment to inform decisions on ESL/EFL students' entry into English-medium university. In P. Davidson, C. Coombe, and W. Jones (Eds.), *Assessment in the Arab World* (pp. 47-59). Dubai: TESOL Arabia.

- Davies, A. (1991). Language Testing in the 1990s. In C. Alderson and B. North, *Language Testing in the 1990s* (pp. 136-149). London and Basingstoke: Macmillan Publishers Ltd.
- Davies, A. (2007). Assessing academic English language proficiency: 40+ years of UK language tests. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner and C. Doe (Eds.), *Language Testing Reconsidered* (pp.73-83). Ottawa: University of Ottawa Press.
- Davies, A. (2008). *Assessing Academic English: Testing English proficiency 1950-1989*. Studies in Language Testing (Vol. 23). Cambridge University Press: UK.
- Davies, A. (2012). Kane, validity and soundness. *Language Testing*, 29(1), 37-42.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., and McNamara, T. (1999). *Dictionary of Language Testing*. Studies in Language Testing (Vol. 7). Cambridge: UCLES/Cambridge University Press.
- Davis, F. (1973). The Martian and the Convert: Ontological Polarities in Social Research. *Journal of Contemporary Ethnography*, 2(3), 333-343.
- Deakin, G. (1996). IELTS in context: issues in EAP for overseas students. *EA Journal*, 15(2), 7-28.
- De Beaugrande, R. (1984). *Text production: Towards a science of composition*. Norwood, NJ: Ablex.
- Denzin, N. (1978). *The Research Act: A theoretical introduction to sociological methods*. New York: McGraw-Hill.

- Denzin, N., and Lincoln, Y. (2005). *The SAGE handbook of qualitative research* (3rd Ed.). Thousand Oaks, CA: Sage.
- Dooley, P. (1999). An investigation into the predictive validity of the IELTS Test as an indicator of future academic success. In K. Martin, N. Stanley and N. Davison (Eds.), *Teaching in the Disciplines/Learning in Context* (pp.114-118). Proceedings of the 8th Annual Teaching Learning Forum, University of Western Australia.
- Douglas, D. (1997). *Testing speaking ability in academic contexts: Theoretical considerations*. TOEFL Monograph Series (Number 8). Princeton: Educational Testing Service.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2010). *Understanding Language Testing*. London: Hodder-Arnold.
- Doye, P. (1991). Authenticity in foreign language testing. In S. Anivan (Ed.), *Current developments in language testing* (pp. 103–10). Singapore: SEAMEO Regional Language Centre.
- Druckman, D. (2005). *Doing Research: Methods of Inquiry for Conflict Analysis*. Thousand Oaks, CA: Sage.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), pp. 155-185.
- Ede, L., and Lunsford, A. (1984). Audience Addressed/Audience Invoked: The Role of Audience in Composition Theory and Pedagogy. *College Composition and Communication*, 35(2), pp. 155-171.

- Engestrom, Y. (1987) *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.
- Ernest, P. (1994). *An Introduction to Research Methodology and Paradigms*. Research Support Unit, School of Education: University of Exeter.
- Everret, R., and Coleman, J. (2003). A critical evaluation of selected IELTS preparation materials. *IELTS Research Reports* (Vol. 5, pp.1-84). Cannebra: Australia.
- Falvey, P., and Shaw, S. (2006). IELTS writing: revising assessment criteria and scales (Phase 5). *Research Notes*, 23, pp. 7-12.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70-85.
- Ferguson, G., and White, E. (1998). A small-scale study of predictive validity. *Melbourne Papers in Language Testing*, 7, 15-63.
- Field, J. (2004). *Psycholinguistics: The Key Concepts*. London: Routledge.
- Flower, L. (1989). Cognition, Context, and Theory Building. *College Composition and Communication*, 40(3), 282-311.
- Flower, L., and Hayes, J. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365-387.
- Frechtling, J., and Sharp, L. (Eds.) (1997). *User-Friendly Handbook for Mixed Methods Evaluation*. Washington, DC: National Science Foundation. Retrieved from [www.ehr.nsf.gov/ehr/rec/pubs/nsf97-153/start.htm](http://www.ehr.nsf.gov/ehr/rec/pubs/nsf97-153/start.htm).

- Friedlander, A. (1990). Composing in English: Effects of a first language on writing in English as a second language. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 109–125). Cambridge: Cambridge University Press.
- Fulcher, G. (2009). *Language Testing Resources*. Retrieved May 7 2009 from <http://www.languagetesting.info>.
- Fulcher, G. (2010). *Rating Scales and the Halo Effect*. Retrieved May 7 2010 from <http://languagetesting.info/features/haloring/rating.html>.
- Fulcher, G., and Davidson, F. (2007). *Language Testing and Assessment*. London and New York: Routledge.
- Furneaux, C., and Rignall, M. (2007). The effect of standardisation-training on rater judgements for the IELTS Writing Module. In L. Taylor and P. Falvey (Eds.), *IELTS Collected Papers: Research in speaking and writing assessment* (Studies in Language Testing, Vol. 19, pp. 422-445). Cambridge: Cambridge University Press/UCLES Cambridge ESOL.
- Garrett, H. (1937). *Statistics in Psychology and Education*. New York: Longmans, Green.
- Geisinger, K. (1992). The metamorphosis in test validity. *Educational Psychologist*, 27(2), 197-222.
- Gerloff, P. (1987). Identifying the unit of analysis in translation: Some uses of think-aloud protocol data. In C. Faerch and C. Kasper (Eds.), *Introspection in second language research* (pp. 135-158). Philadelphia: Multilingual Matters.
- Golder, K., Reeder, K., and Sarah, F. (2009). Determination of appropriate IELTS band score for admission into a programme at a Canadian post-secondary poly-technic

institution. *IELTS Research Reports*, (Vol. 10, pp. 69-93). IDP: IELTS Australia and British Council.

Grabe, W., and Kaplan, R. B. (1996). *Theory and Practice of Writing: An Applied Linguistic Perspective*. London: Longman.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. (Studies in Language Testing Series, Vol. 5). Cambridge: University of Cambridge Local Examinations Syndicate/ Cambridge University Press.

Green, A. (2007a). *IELTS washback in context: preparation for academic writing in higher education*. (Studies in Language Testing, Vol. 25). Cambridge: Cambridge ESOL/ Cambridge University Press.

Green, A. (2007b). Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97.

Hammersley, M., and Atkinson, P. (2007). *Ethnography: principles in practice* (3rd ed.). London: Routledge.

Hamp-Lyons, L. (1986). *Testing Second Language Writing in Academic Settings*. Unpublished Doctoral Dissertation. University of Edinburgh.

Hamp-Lyons, L. (1991a). The Writer's Knowledge and our Knowledge of the Writer. In Hamp-Lyons, L. (Ed.), *Assessing second language writing in academic contexts* (pp. 241-276). Norwood, NJ: Ablex Publishing.

Hamp-Lyons, L. (Ed.). (1991b). *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex Publishing.

- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, 8(1), pp. 5–16.
- Hamp-Lyons, L., and B. Kroll (1997). *TOEFL 2000 - Writing: Composition, Community, and Assessment* (TOEFL Monograph Series 5). Princeton, NJ: Educational Testing Service.
- Hardman, J. (2007). Making Sense of the Meaning Maker: Tracking the object of activity in a computer-based mathematics lesson using activity theory. *International Journal of Education and Development using ICT*, 3(4). Retrieved from <http://ijedict.dec.uwi.edu/viewarticle.php?id=423>.
- Hatch, E., and Farhady, H. (1982). *Research Design and Statistics for Applied Linguists*. Rowley, Massachusetts: Newbury House.
- Hayes, B., and Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97-112). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In C. Levy and S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah: New Jersey: Lawrence Erlbaum Associates.
- Hayes, J., and Flower, L. (1980). Identifying the Organization of Writing Processes. In: L. Gregg and E. Steinberg (Eds.), *Cognitive Processes in Writing: An Interdisciplinary Approach* (pp.31-50). Hillsdale, NJ: Lawrence Erlbaum.
- Hawkey, R. (2004). An IELTS impact study: Implementation and some early findings. *Research Notes*, 15. Cambridge: Cambridge ESOL.

- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. (Studies in Language Testing, Vol. 24, pp. 1-235). Cambridge: Cambridge University Press.
- Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House.
- Hill, K., Storch, N., and Lynch, B. (1999). A Comparison of IELTS and TOEFL as Predictors of Academic Success. *Research Reports* , 2, 52-63.
- Hodkinson, P., and Smith, K. J. (2008). The Relationship between Research, Policy and Practice. In G. Thomas and R. Pring (Eds.), *Evidence-Based Practice in Education*. Maidenhead: Open University Press.
- Horowitz, R. (1986). Remaining an Outsider: Membership as a Threat to Research Rapport. *Journal of Contemporary Ethnography*, 14(4), 409-430.
- Hout, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237–263.
- Hughes, A. (1994). *Backwash and TOEFL 2000*. Unpublished manuscript, commissioned by Educational Testing Service (ETS). University of Reading, UK.
- Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hyatt, D., and G. Brooks (2009). Investigating Stakeholders' Perceptions of IELTS as an entry requirement for higher education in the UK. . *IELTS Research Reports*, (Vol. 10, pp. 17-68). IELTS Australia: Canberra and British Council: London.

- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *ESP Journal*, 13(3), 239-56.
- Hyland, K. (2002). *Teaching and researching writing*. Applied Linguistics in action series. London: Longman.
- Hyland, K. (2003). Genre-based pedagogies: A social response to process. *Journal of Second Language Writing*, 12, 17-29.
- Hymes, D. (1966). On communicative competence. Paper originally read at the Research Planning Conference on Language Development among Disadvantaged Children, Yeshiva University, June 1966. Reprinted, in part, in C.J. Brumfit and K. Johnson (Eds.). (1979). *The Communicative Approach to Teaching*. Oxford: Oxford University Press.
- IELTS Revision Specifications (1993). Draft version 7. Cambridge: UCLES.
- IELTS Exam Preparation (2007). *Sample IELTS Writing Tasks*. Retrieved October 30 2012 from <http://www.ielts-exam.net>.
- IELTS (2007). *The IELTS Handbook*. Cambridge: UCLES, The British Council, IDP Australia.
- IELTS (2009). *Information for Candidates*. Retrieved June 8 2009 from [http://www.ielts.org/PDF/Information\\_for\\_Candidates\\_2007.pdf](http://www.ielts.org/PDF/Information_for_Candidates_2007.pdf).
- IELTS (2010a). *Score processing, reporting and interpretation*. Retrieved June 3 2010 from [http://www.ielts.org/researchers/score\\_processing\\_and\\_reporting.aspx](http://www.ielts.org/researchers/score_processing_and_reporting.aspx).
- IELTS (2010b). *Test results*. Retrieved June 3 2010 from [http://www.ielts.org/test\\_takers\\_information/faqs/test\\_results.aspx](http://www.ielts.org/test_takers_information/faqs/test_results.aspx).

- IELTS (2011). *IELTS Application Form*. Retrieved December 13 2012 from [http://www.ielts.org/PDF/IELTS\\_Application\\_Form\\_November2011.pdf](http://www.ielts.org/PDF/IELTS_Application_Form_November2011.pdf).
- IELTS (2012a). *USA says yes to IELTS*. Retrieved October 16 2012 from <http://www.ielts.org/usa/>.
- IELTS (2012b). *History of IELTS*. Retrieved October 17 2012 from [http://www.ielts.org/researchers/history\\_of\\_ielts.aspx](http://www.ielts.org/researchers/history_of_ielts.aspx).
- IELTS (2012c). *Who owns IELTS*. Retrieved October 17 2012 from [http://www.ielts.org/institutions/about\\_ielts/who\\_owns\\_ielts.aspx](http://www.ielts.org/institutions/about_ielts/who_owns_ielts.aspx).
- IELTS (2012d). *IELTS Jointly Funded and Published Research*. Retrieved October 16 2012 from <http://www.ielts.org/researchers/research.aspx>.
- IELTS (2012e). *IELTS Guide for Teachers*. Retrieved December 7 2012 from [http://www.ielts.org/PDF/IELTS%20Guide%20for%20Teachers\\_British\\_FA%202\\_LR\(v.2\).pdf](http://www.ielts.org/PDF/IELTS%20Guide%20for%20Teachers_British_FA%202_LR(v.2).pdf).
- IELTS (2012f). *Examiner information*. Retrieved April 26 2012 from [http://www.ielts.org/researchers/examiner\\_information.aspx](http://www.ielts.org/researchers/examiner_information.aspx).
- Ingram, D., and Bayliss, A. (2007). IELTS as a predictor of academic language performance, Part 1. *IELTS Research Reports*, (Vol. 7, pp.137-204). IELTS Australia: Canberra and British Council: London.
- Johns, A. (1986). Coherence and academic writing: Some definitions and suggestions for teaching. *TESOL Quarterly*, 20(2), pp. 247-265.
- Johnson, R., and Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33 (7), 14-26.

- Jones, S., and Tetroe, J. (1987). Composing in a second language. In A. Matsuhashi (Ed.), *Writing in real time: Modeling production processes* (pp. 34–57). Norwood, NJ: Ablex.
- Jorgensen, D. (1989). *Participant Observation: A Methodology for Human Studies*. Newbury Park, CA: Sage Publications.
- Kaplan, B., and Maxwell J. (1994). Qualitative Research Methods for Evaluating Computer Information Systems. In J. Anderson, C. Aydin and S. Jay (Eds.), *Evaluating Health Care Information Systems: Methods and Applications* (pp.45-68). Thousand Oaks, CA: Sage.
- Kellogg, R. (1996). A Model of Working Memory in Writing. In C. Levy and S. Ransdell (Eds.), *The Science of Writing*. Mahwah, NJ: Lawrence Erlbaum.
- Kerstjens, M., and Nery, C. (2000). Predictive validity in the IELTS Test: A study of the relationship between IELTS scores and students' subsequent academic performance. In R. Tulloh (Ed.), *IELTS Research Reports*, (Vol. 3, pp. 85 -108). Canberra: IELTS Australia.
- King Faisal University (2012). *The College of Arts*. Retrieved October 15 2012 from <http://www.kfu.edu.sa/ar/Colleges/arts/departments/Pages/home.aspx#>.
- Kobayashi, H., and Rinnert C. (1994). Effects of first language on second language writing: Translation versus direct composition. In A. Cumming (Ed.), *Bilingual performance in reading and writing* (pp. 223- 255). Ann Arbor, MI: Research Club in Language Learning.
- Krapels, A. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37–56). Cambridge: Cambridge University Press.

- Krings, H. (1987). The use of introspection data in translation. In C. Faerch, and G. Kasper, *Introspection in second language research* (pp. 159–176). Clevedon: Multilingual Matters.
- Krings, H. (1989). Writing in a foreign language: Process analyses of the fourth skill. In G. Antos and H. Krings, *Textproduktion. Ein interdisziplinärer Forschungsüberblick* (pp. 377- 436). Tübingen: Niemeyer.
- Krings, H. (1992). Empirical studies of foreign language writing processes : A research survey. In W. Börner, and K. Vogel, *Schreiben in der Fremdsprache. Prozeß und Text, Lehren und Lernen* (pp.47–77). Bochum: AKS-Verlag.
- Kroll, B., and Reid, J. (1994). Guidelines for designing writing prompts: clarifications, caveats and cautions. *Journal of Second Language Writing*. 3(3), 231-255.
- Lankshear, C., and Knobel, M. (2004). *A Handbook for Teacher Research: From design to Implementation*. Bell and Bain Ltd: Glasgow.
- Lantolf, J. (2000). Introducing sociocultural theory. In J. Lantolf (Ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press.
- Lantolf, J., and Thorne, S. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.
- Lay, N. (1982). Composing processes of adult ESL learners: A case study. *TESOL Quarterly*, 16, 406-411.
- Lewkowicz, J. (2000). Authenticity in language testing: some outstanding questions. *Language Testing*, 17(43), 43-64.

- Libarkin, J., and Kurdziel, J. (2002). Research Methodologies in Science Education: The Qualitative/Quantitative Debate. *Journal of Geoscience Education*, 50, 78-86.
- Liebman-Kline, J. (1986). Toward a contrastive rhetoric: A rhetoric of process. Paper presented at the Annual Meeting of Teachers of English to Speakers of Other languages, Anaheim, CA. (ERIC Document Reproduction Service No. ED 271963).
- Lincoln, Y., and Guba, E. (1985). *Naturalistic inquiry*. New York: Sage.
- Lloyd, D., Davidson, P., and Coombe, C. (Eds.). (2005). *The Fundamentals of Language Assessment: A Practical Guide for Teachers in the Gulf*. Dubai: TESOL Arabia.
- Lumley, T. (2002). Assessment criteria in a large scale writing test: what do they really mean to the raters? *Language Testing*, 19(3), 246-276.
- Luoma, S. (2001). *What Does Your Language Test Measure?* Unpublished PhD Dissertation. University of Jyväskylä, Jyväskylä.
- Lynch, A. (1982). Authenticity in language teaching: some implications for the design of listening materials. *British Journal of Language Teaching*, 20(1), 9–16.
- Manchon, R., Roca de Larios, J., and Murphy, L. (2000). An approximation to the study of backtracking in L2 writing. *Learning and Instruction*, 10(1), 13–35.
- Marsh, H., and Ireland, R. (1987). The Assessment of Writing Effectiveness: A Multidimensional Perspective. *Australian Journal of Psychology*, 39(3), 353-367.
- Marshall, M. (1996). Sampling for qualitative research. *Family Practice*, 13(6), 522-525.
- Matsumoto, K. (1995). Research Paper Writing Strategies of Professional Japanese EFL Writers. *TESL Canada Journal*, 13(1), 17-27.

- Mayor, B., Hewings, A., North, S., Swan, J., and Coffin, C. (2007). A Linguistic analysis of Chinese and Greek L1 scripts for IELTS Academic Writing Task2. In L. Taylor and P. Falvey (Eds.), *IELTS collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 250-315). Cambridge: Cambridge University Press.
- Mayor, B., Hewings, A., and Swann, J. (2003). *An Analysis of the Linguistic Features of Output from IELTS Academic Writing 1 and 2*. London: British Council/IDP Australia.
- McDowall, C., and Merrylees, B. (1998). Survey of receiving institutions' use and attitude to IELTS. *IELTS Research Reports*, (Vol. 1, pp. 116-139). IELTS Australia, Canberra.
- McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11(2), 113–128.
- McNamara, T. (1995). Modelling performance: opening Pandora's Box. *Applied Linguistics*, 16(2), 159–75.
- McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.
- McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance. *Applied Linguistics*, 18(4), 446-466.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- McNamara, T. (2006). Language Testing. In A. Davies and C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 763-783). Oxford: Basil Blackwell.
- McRoy, R. (no date). *Qualitative Research*. Retrieved April 13 2009 from [http://www.uncp.edu/home/marson/qualitative\\_research.html](http://www.uncp.edu/home/marson/qualitative_research.html).

- Merriam, S. (1998). *Qualitative Research and Case Study: Applications in Education*. San Francisco, CA: Jossey-Bass.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement*. (3rd Ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1996a). *Standards-based score interpretation: Establishing valid grounds for valid inferences*. Proceedings of the joint conference on standard setting for large scale assessments, Sponsored by National Assessment Governing Board and The National Center for Education Statistics. Washington, DC: Government Printing Office.
- Messick, S. (1996b). Validity of Performance Assessment. In G. Philips (1996). *Technical Issues in Large-Scale Performance Assessment*. Washington, DC: National Center for Educational Statistics.
- Mickan, P. (2003). 'What's your score?' An investigation into language descriptors for rating written performance. In L. Hyam (Ed.), *IELTS Research Reports*, (Vol. 5, pp. 125–157). Canberra: IELTS Australia Pty Ltd.
- Mickan, P., and Motteram, J. (2008). An ethnographic study of classroom instruction in an IELTS preparation program. *IELTS Research Reports*, (Vol. 8, pp. 17-43). Canberra: IELTS Australia Pty Ltd.
- Mickan, P., and Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. *IELTS Research Reports*, (Vol. 10, pp. 223-262). IELTS Australia: Canberra and British Council: London.

- Mickan, P., and Slater, S. (2003). Text analysis and the assessment of academic writing. In R. Tulloh (Ed.), *IELTS Research Reports*, (Vol. 4, pp. 59–88). Canberra: IELTS Australia Pty.
- Mickan, P., Slater, S., and Gibson, C. (2000). A study of response validity of the IELTS writing subtest. *IELTS Research Reports*, (Vol. 3, pp. 29-48). Canberra, Australia: IDP IELTS Australia.
- Miles, M., and Huberman, M. (1984). *Qualitative data analysis*. Beverly Hills, CA: Sage.
- Ministry of Education (2012). *Saudi Arabia Educational Policy*. Retrieved December 12 2012 from: <http://www.moe.gov.sa/Pages/educationPolicy.aspx>.
- Ministry of Higher Education (2012). *Education in the Kingdom*. Retrieved October 3 2012 from <http://www.mohe.gov.sa/ar/studyinside/Pages/default.aspx>.
- Moore, T., and Morton, J. (2007). Authenticity in the IELTS Academic Module Writing test: a comparative study of Task 2 items and university assignments. In L. Taylor and P. Falvey (Ed.), *IELTS collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 197-249). Cambridge: Cambridge University Press.
- Moragne e Silva, M. (1988). A study of composing in a first and second language. *Texas Papers in Foreign Language Education*, 1(2), 132-151.
- Morrow, K. (1991). Evaluating communicative tests. In S. Anivan, (Ed.), *Current developments in language testing*, (pp. 111–18). Singapore: SEAMEO Regional Language Centre.

- Mu, C. (2005). A Taxonomy of ESL Writing Strategies. In *Redesigning Pedagogy: Research, Policy, Practice*. (Unpublished Manuscript). Retrieved March 18 2010 from <http://eprints.qut.edu.au/64/>.
- Nisbet, J., and Watt, J. (1984). Case Study. In J. Bell, T. Bush, A. Fox, J. Goodey and S. Goulding (Eds.), *Conducting Small-scale Investigations in Educational Management* (pp.79-92). London: Harper and Row.
- Nystrand, M. (1989). A social-interactive model of writing. *Written Communication*, 6(1), 66-85.
- Odell, L. (1981). Defining and assessing competence in writing. In C. Cooper (Ed.), *The nature and measurement of competency in English* (pp. 95 - 138). Urbana, IL: National Council of Teachers of English.
- O'Loughlin, K. (2008). The use of IELTS for university selection in Australia: A case study. *IELTS Research Reports*, (Vol. 8, pp.145-241). Canberra: IELTS Australia Pty Ltd.
- O'Loughlin, K., and Wigglesworth, G. (2007). Investigating task design in Academic Writing Prompts. In L. Taylor and P. Falvey (Ed.), *IELTS collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 379-421). Cambridge: Cambridge University Press.
- Onwuegbuzie, A., and Johnson, R. (2006). The validity issues in mixed research. *Research in Schools*, 13(1), 48-63.
- Oppenheim, A. (2000). *Questionnaire Design, Interviewing and Attitude Measurement*. London: Continuum.

- O'Sullivan, B. (2000). *Towards a Model of Performance in Oral Language Testing*. Unpublished PhD Dissertation. University of Reading.
- O'Sullivan, B., and Rignall, M. (2007) Assessing the value of bias analysis feedback to raters for the IELTS Writing Module. In L. Taylor and P. Falvey (Ed.), *IELTS collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 446-478). Cambridge: Cambridge University Press.
- Palys, T. (1997). *Research decisions: Quantitative and qualitative perspectives* (2nd Ed.). Toronto, ON: Harcourt Brace Jovanovich.
- Papajohn, D. (1999). The effect of topic variation in performance testing: the case of the chemistry TEACH test for international teaching assistants. *Language Testing*, 16(1), 52-81.
- Philips, D. (1987). Validity in qualitative research: Why the worry about warrant will not wane. *Education and Urban Society*, 20(1), 9-24.
- Picard, M. (2007). English entrance requirements and language support for international postgraduate students. In *Enhancing Higher Education, Theory and Scholarship* (pp. 439-449). Proceedings of the 30th HERDSA Annual Conference, Adelaide.
- Poehner, M. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal*, 91(3), 323-340.
- Pring, R. (2004). *Philosophy of Educational Research*. London: Continuum.
- Prior, P. (2006). A sociocultural theory of writing. In C. MacArthur, S. Graham, and J. Fitzgerald (Eds.), *Handbook of Writing Research*, (pp. 54–66). New York: The Guilford Press.

- Qi, D. (1998). An inquiry into language-switching in second language composing processes. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 54, 413–435.
- Raimes, A. (1985). What unskilled writers do as they write. *TESOL Quarterly*, 19(2), 229–258.
- Raimes, A. (1987). Language proficiency, writing ability, and composition strategies: A study of ESL college student writers. *Language Learning*, 37(3), 439–467.
- Rao, C., McPherson, K., Chand, R., and Khan, V. (2003) Assessing the impact of IELTS preparation programs on candidates' performance on the General Training Reading and Writing test modules. *IELTS Research Reports*, (Vol. 5, pp. 237-262). Canberra: IELTS Australia Pty Ltd.
- Rea-Dickins, P., Kiely, R., and Yu, G. (2007). Student Identity, Learning and Progression (SILP): The affective and academic impact of IELTS on 'successful' candidates. In P. McGovern and S. Walsh (Eds.), *IELTS Research Reports*, (Vol. 7, pp. 59-136). IELTS Australia: Canberra and British Council: London.
- Rea-Dickins, P., Kiely, R., and Yu, G. (2011). Uses and impact of test scores in university admissions process: The language test as the "hard" criterion. In B. O'Sullivan (Ed.), *Language testing: Theories and practices*, (pp.262-281). Basingstoke: Palgrave Macmillan.
- Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, 9(2), 109-121.
- Rignall, M., and Furneaux, C. (2007). *The effect of standardisation training on rater judgements for the IELTS Writing Module*. In L. Taylor and P. Falvey (Ed.), *IELTS*

*collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 422-445). Cambridge: Cambridge University Press.

Robson, C. (2002). *Real world research* (2nd Ed.). Oxford: Blackwell.

Roca de Larios, J., Manchón, R., Murphy, L., and Marín, J. (2008). The foreign language writers' strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing*, 17(1), 30–47.

Roca de Larios, J., Murphy, L., and Mancho'n, R. (1999). The use of restructuring strategies in EFL writing: A study of Spanish learners of English as a foreign language. *Journal of Second Language Writing*, 8(1), 13–44.

Rohman, D. (1965). Pre-Writing: The Stage of Discovery in the Writing Process. *College Composition and Communication*, 16, 106-12.

Ruth, L., and Murphy, S. (1984). Designing Topics for Writing Assessment: Problems of Meaning. *College Composition and Communication*, 35(4), 410-421.

Ruth, L., and Murphy, S. (1988). *Designing Tasks for the Assessment of Writing*. Norwood, NJ: Ablex.

Sakyi, A. (2000). Validation of holistic scoring for ESL writing assessment: How raters evaluate compositions. In J. Kunnan, (Ed.), *Fairness and validation in language assessment*. (Studies in Language Testing, Vol. 9, pp. 129–152). Cambridge: Cambridge University Press.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9(3), 259–291.

- Saville, N. (2000). Investigating the Impact of International Language Examinations. *Research Notes*, 2. Cambridge: Cambridge ESOL.
- Sawyer, W. and Singh, M. (2011). Learning to play the ‘classroom tennis’ well: IELTS and international students in teacher education. *IELTS Research Reports*, (Vol. 11, pp.22-73). IELTS Australia: Canberra and British Council: London.
- Schütz, A. (1962). *Collected Papers Vol.1 The problem of social reality*. The Hague: Martinus Nijhoff.
- Shaw, S. (2002a). IELTS Writing: Revising assessment criteria and scales (Phase1). *Research Notes*, 9, 16-18.
- Shaw, S. (2002b). IELTS Writing: Revising assessment criteria and scales (Phase2). *Research Notes*, 10, 10-13.
- Shaw, S. (2003a). *IELTS Writing Assessment Revision Project (Phase3): Validating the revised rating scale – A quantitative analysis*. Cambridge: UCLES internal report.
- Shaw, S. (2003b). IELTS Writing: Revising assessment criteria and scales (Phase3). *Research Notes*, 16, 3-7.
- Shaw, S. (2004). IELTS Writing: Revising assessment criteria and scales (Concluding Phase2). *Research Notes*, 15, 9-11.
- Shaw, S. (2005). *IELTS Writing Assessment Revision Working Group: Summary of progress (June, 2001- October, 2005)*. Cambridge: UCLES internal report.
- Shaw, S. and Falvey, P. (2008). The IELTS Writing Assessment Revision Project: Towards a revised rating scale. *Research Reports*, 1.

- Shaw, S., and Weir, C. (2007). *Examining Writing: Research and Practice in Assessing Second Language Writing*. (Studies in Language Testing, Vol. 26). Cambridge: UCLES/CUP.
- Shohamy, E. (2001). *The Power of Tests: A critical perspective on the uses of language tests*. Harlow: Longman.
- Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11–23). New York: Cambridge University Press.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4), 657-77.
- Skehan, P. (1991). Progress in Language Testing: the 1990s. In C. Alderson and B. North, *Language Testing in the 1990s*. (p.3-21). London and Basingstoke: Macmillan Publishers Ltd.
- Smith, H., and Haslett, S. (2007). Attitudes of tertiary key decision-makers towards English language tests in Aotearoa New Zealand: Report on the results of a national provider survey. *IELTS Research Reports*, (Vol. 7, pp. 13-57). IELTS Australia: Canberra and British Council: London.
- Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31(4), 378–388.
- Spack, R. (1988). Initiating ESL Students into the Academic Discourse Community: How far should we go? *TESOL Quarterly*, 22(1), 29–51.

- Spack, R. (1997). The acquisition of academic literacy in a second language: A longitudinal case study. *Written Communication*, 14(1), 3–62.
- Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, 18(4), 464-481.
- Spolsky, B. (1985). The limits of authenticity in language testing. *Language Testing*, 2(1), 31–40.
- Spolsky, B. (1998). What is the user's perspective in language testing? Paper presented at the Colloquium, *The State Of The Art In Language Testing: The User's Perspective*. National Foreign Language Centre. The John Hopkins University, Washington DC.
- Spolsky, B. (2008). Introduction: Language Testing at 25: Maturity and responsibility? *Language Testing*, 25(3), 297–305.
- Starfield, S. (2004). Word power: negotiating success in a first-year sociology essay. In L. Ravelli and R. Ellis (Eds.), *Analyzing academic writing: Contextualised frameworks* (pp. 66-83). London: Continuum.
- Sullivan, F. (1997). Calling writers' bluffs: The social production of writing ability in university placement testing. *Assessing Writing*, 4(1), 53-81.
- Taylor, L. (2007). The impact of the joint-funded research studies on the IELTS Writing Module. In L. Taylor and P. Falvey (Ed.), *IELTS collected papers: research in speaking and writing*. (Studies in Language Testing, Vol. 19, pp. 479-492). Cambridge: Cambridge University Press.
- Tillema, M., Bergh, H., Rijlaarsdam, G., and Sanders, T. (2013). Quantifying the quality difference between L1 and L2 essays: A rating procedure with bilingual raters and

L1 and L2 benchmark essays. *Language Testing*, 30(1), 71-97.

The National Centre for Assessment in Higher Education (2012). *Exams*. Retrieved October 4 2012 from <http://www.qeyas.com/Qiyas/Exams/Exams.aspx>.

The Saudi Arabian Cultural Mission to the US (2012). *The educational ladder in Saudi Arabia*. Retrieved October 4 2012 from [http://www.sacm.org/Publications/58285\\_Edu\\_complete.pdf](http://www.sacm.org/Publications/58285_Edu_complete.pdf).

Uysal, H. (2010). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314-320.

Uzawa, K. (1996). Second language learners' processes of L1 writing, L2 writing and translation from L1 into L2. *Journal of Second Language Writing*, 5(3), 271–294.

Uzawa, K., and Cumming, A. (1989). Writing strategies in Japanese as a foreign language: Lowering or keeping up the standards. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 46, 178–194.

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., and Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. *Journal of Second Language Writing*, 18(4), 235–250.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes* (M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, Eds.) Cambridge, MA: Harvard University Press.

Walford, G. (2001). *Doing qualitative educational research*. London: Continuum.

- Wall, D., Calpham, C. and Alderson, C. (1999). Validating Tests in Difficult Circumstances. In C. Alderson and B. North, *Language Testing in the 1990s*. (pp. 209-225). Macmillan Publishers Ltd: London and Basingstoke.
- Wang, L. (2003). Switching to first language among writers with differing second-language proficiency. *Journal of Second Language Writing*, 12(1), 347–375.
- Wang, W., and Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing*, 11(3), 225–246.
- Webber, P. (1990). *Basic Content Analysis* (2<sup>nd</sup> edition). Thousand Oaks, CA: Sage.
- Weigle, S. (1994). Effects of training on raters of ESL compositions. *Language Testing*, 11(2), 197–223.
- Weigle, S. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263–87.
- Weigle, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weir, C. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- Weir, C., Hawkey, R., Green, A., Unaldi, A. and Devi, S. (2009). The relationship between the academic reading construct as measured by the IELTS and the reading experiences of students in their first year of study at a British university. *IELTS Research Reports*, (Vol. 9, pp. 97-156). IELTS Australia: Canberra and British Council: London.

- Weissberg, B. (2000). Developmental relationships in the acquisition of English syntax: Writing vs. speech, *Learning and Instruction*, 10(1), 37-53.
- Wertsch, J. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.
- Wette, R. (2012). How Professionally Relevant Can Language Tests Be?: The Author Responds. *Language Assessment Quarterly*, 9(1), 109-112.
- Whalen, K., and Menard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, 45(3), 381-418.
- White, E. (1985). *Teaching and Assessing Writing*. Jossey-Bass: San Francisco.
- Wolfe, E., Kao, C., and Ranney, M. (1998). Cognitive differences in proficient and non proficient essay scorers. *Written Communication*, 15, 465-492.
- Wolfersberger, M. (2003). L1 to L2 writing process and strategy transfer: A look at lower proficiency writers. *TESL-EJ: Teaching English as a Second or Foreign Language*, 7(2), 1-15.
- Wood, R. (1993). *Assessment and testing*. Cambridge: Cambridge University Press.
- Woodall, B. (2002). Language-switching: Using the first language while writing in a second. *Journal of Second Language Writing*, 11, 7-28.
- Woodrow, L. (2006). Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL*, 1, 51-70.

- Xiaoyue, B. (2010). *The effects of topic familiarity and strategic planning in topic-based task performance at different proficiency levels*. Unpublished PhD dissertation. The Chinese University of Hong Kong.
- Yin, R. (2009). *Case study research: Design and methods* (4th Ed.). Thousand Oaks, CA: Sage.
- Yu, G., Rea-Dickins, P., and Kiely, R. (2011). The cognitive processes of taking IELTS academic writing task one. *IELTS Research Reports*, (Vol. 11, pp. 373 – 449). IELTS Australia: Canberra and British Council: London.
- Zamel, V. (1983). The composing processes of advanced ESL students: Six case studies. *TESOL Quarterly*, 17, 165 - 187.
- Zimmerman, R. (2000). L2 writing: Sub-processes, a model of formulating and empirical findings, *Learning and Instruction*, 10(1), 73 - 99.

# APPENDICES

## APPENDIX I

### DATA COLLECTION TOOLS & KEY FINDINGS OF THE STUDY

#### Phase I

#### IELTS Awareness Form for Candidates (Pre-Writing Test Phase I)

Dear Respondent

I am conducting this study for the purpose of investigating the validity (does the test measure what it intends to measure) of the Academic Writing Module (AWM) in the IELTS test. Your answers will help me to know about your opinion of the validity of the test and how you make sense of its academic writing tasks.

My aim is to put the information of this questionnaire together with other data to understand why do students fail to cope with the demands of the university despite the fact they have met the minimum entry requirements through IELTS. Your answers are highly confidential and you have the right to withdraw from the study whenever you like.

#### Part1: Personal Information

1- Age \_\_\_\_\_ 2- Sex: Male  Female  3- Level of education

(now):

Undergraduate  Postgraduate  4- Course \_\_\_\_\_

5- Year of study \_\_\_\_\_

#### Part2: Your prospective course and place of study

6- Are you planning to study in: UK  US  Canada  Australia

other: \_\_\_\_\_

7- What subject and level do you intend to study?

-Subject: \_\_\_\_\_ -Level: Undergraduate  Postgraduate

#### Part3: About the AWM in the IELTS test

8- I am taking the IELTS test because:

---

---

9- I need to achieve an overall score of \_\_\_\_\_ with a sub-score of \_\_\_\_\_ in writing.

10- The writing test lasts for \_\_\_\_\_ minutes.

11- At the beginning of the writing test, I will:

---

---

12- Task 1 is different from Task 2 in terms of:

---

---

13- My answers in both tasks should be:

---

---

14- I should understand the question otherwise:

---

---

15- I should write \_\_\_\_\_ words in Task 1.

16- While writing, I should pay attention to:

---

---

17- The allotted time for Task 2 is \_\_\_\_\_ minutes.

18- The strategy/ies that I am going to use to succeed in the writing test is/are:

---

---

19- What challenges do expect to face in the writing test? And why?

---

---

*Thank you very much for filling out this questionnaire* 🍷

**IELTS Awareness Form for Candidates (Post-Writing Test Phase I)**

Dear Respondent

Now, I would like to know more information about how did you go through Task 1 & Task 2.

**Questions about Task 1 & Task 2**

1- What do you think of the time given to complete the tasks? And how much time did you spend on each task?

---

---

2- How important are the guidelines/instructions at the beginning of each task?

---

---

3- What did you do before you started to write?

---

---

4- How difficult were the questions of the tasks? Mention the difficult words in the questions that you could not understand?

---

---

5- While writing, how often did you go back to correct grammar mistakes, style and organisation? And with which task you did this more than the other?

---

---

6- How useful is the graph/chart/table in Task 1?

---

---

7- How familiar you were with the content of the tasks?

---

---

8- The good thing/s about the IELTS writing task is/are:

---

---

9- The problem/s with IELTS is/are:

---

---

*Thank you very much for filling out this questionnaire 🙌*

### **IELTS Post-Writing Test Interview (Candidates)**

1-Is this your first IELTS test? If not, how many times have you taken the test, when and where?

2-In your opinion, was there enough time to make drafts and complete each task?

If not, which task and why?

3-Do you think it is important to have clear guidelines at the beginning of each task?

If yes, which task you think should be clearer?

4- Did you brainstorm (think deeply to organise your ideas) before starting to write? If not, what did you do?

5-What strategies did you use while writing?

6-In which task you exerted more effort? And why?

7- Does it matter which writing task you start with? If yes, what is the difference?

8-Did you read the question carefully to make sure you understood what was required?

9-Did the question help you in writing relevant information? How often did you go back to it?

10- Do you think the more detailed the question of the composition task the more ideas you will have?

If yes, did you try to paraphrase key words in the question and include them in your answer?

11-Was it useful to go back to correct grammar mistakes, style and organisation? If not, please explain?

12-While writing, did you relate the topic of the task to what you already know? If yes, in which task you did this frequently?

13-Is there anything you would like to add about IELTS academic writing tasks to help inform my research?

## **IELTS Questionnaire for Markers**

Dear Respondent

I am conducting this study for the purpose of investigating the validity (does the test measure what it intends to measure) of the Academic Writing Module (AWM) in the IELTS test. Your answers will help me to know about your opinion of the scoring validity of the test and the extent to which the rating scales are used consistently while scoring academic writing tasks.

My aim is to put the information of this questionnaire together with other data to understand why do students fail to cope with the demands of the university despite the fact they have met the minimum entry requirements through IELTS. Your answers are highly confidential and you have the right to withdraw from the study whenever you like.

### **Part 1: Personal Information**

- 1- Age \_\_\_\_\_.
- 2- Gender: Male  Female
- 3- Level of education: Undergraduate  Postgraduate  4-Major \_\_\_\_\_.
- 5- First Language \_\_\_\_\_.
- 6- Nationality \_\_\_\_\_.
- 7- Relevant experience and training \_\_\_\_\_.
- 8- Years of employment in the British Council \_\_\_\_\_.

### **Part 2: About the Scoring of AWM in the IELTS Test**

9- How often do you refer to the scoring guide?

---

---

10- How important are the training sessions for scoring?

---

---

11-Do you think some candidates are awarded different scores in different centres or occasions? Why?

---

---

12- What is the difference between the scoring of Task 1 and Task 2 in the AWM?

---

---

13- What are the features of the text that you look for while scoring both tasks?

Task1 \_\_\_\_\_

Task2 \_\_\_\_\_

14- What do you think of the validity of the scoring method used in the AWM?

---

---

15- Do you think single marking is adequate? If not, how often do you ask for double marking?

---

---

16- What do you think of the clarity of the criteria for scoring AWM?

---

---

17- After looking at the criteria of the candidate's text separately, how do you put these together to get the overall score?

---

---

18- Which factors interfere most with the consistency of scoring the AWM?

---

---

19- Is there anything else you would like to add?

---

---

*Thank you very much for your participation*

## Phase II

### IELTS Tasks and University Tasks Form for Candidates (Phase II)

Dear Respondent

In this questionnaire, I would like to know about the authenticity/correspondence or the relationship between the IELTS writing test tasks and the university tasks.

#### Questions

1- What was your expectation of the university tasks when you first sat for the IELTS test?

---

---

2- In what ways IELTS writing tasks are different from university tasks?

---

---

3- In what ways IELTS writing tasks are similar to university tasks?

---

---

4- What do you think of the suitability of the IELTS writing tasks for university students?

---

---

*Thank you very much for filling out this questionnaire 🙌*

### **Interview Questions for Candidates**

- 1- What is the relationship between IELTS writing tasks and university tasks?  
(Relevance)
  
- 2- To what extent did the IELTS writing tasks help you in the preparation for university tasks?
  
- 3- Do you consider both of the writing tasks to be representative of university tasks?  
Why do you think so? (Representativeness)
  
- 4- Is there anything you would like to say?

### **Interview Questions for Lecturers**

1- As a lecturer in this school you have probably met so many international students, can I ask you what is your expectation of international students enrolled in your course? And why do you think some students fail to cope with the demands of the university?

2- As you may know international students whose English is not the first language must show evidence that they have adequate English to start their academic courses, now what do you think of the level of proficiency of those students who have satisfied the minimum entry requirements in university?

3- Do you think they are ready to start their courses or need more support to survive at university?

4- When it comes to developing academic writing skills, what would you recommend teachers focus on with L2 writers?

5- Advanced writing especially at university level requires students to produce complex structures and formal language, the assessment process on the part of the teachers then perhaps becomes equally demanding, Could you please tell me if there are any differences or similarities between your assessment criteria for assignments in your course and IELTS criteria? Do students understand them?

6- Can you mention the essential factors that would help students to succeed in your course?

7- Is there anything you would to say?

## Sample of a transcription of one interview

### Lecturers' Perception of the IELTS Test Interview

// Interruptions

... Inaudible

xxx Pause

xxxxxx Long pause

- 1 1) Researcher: Good afternoon Mr. John and Welcome to this interview.
- 2 2) Interviewee: My pleasure.
- 3 3) Researcher: (Giving him the IELTS tasks) Could you please tell me what do you think of these
- 4 tasks?
- 5 4) Interviewee: I think IELTS task one is a sensible task xxx and xxx because it xxx requires students to
- 6 have the English vocabulary and sentence structure to interpret the task around into a text
- 7 5) Researcher: Yeah
- 8 6) Interviewee: And xxx em it also encourages them to think about the presentation data which is about
- 9 a separate issue really and xxxxxxxx I am not sure that describing the histogram is something I would
- 10 expect students to do routinely as part of engagement with research data or as part of the
- 11 assignments//
- 12 7) Researcher: //Well//
- 13 8) Interviewee: //em because I would often encourage people to use the histogram as the statement and
- 14 then the text that would follow would be a discussion of the issues that are raised by that xxx and xxx
- 15 so it might not be a task which is very close to the kinds of tasks that people do when they come to
- 16 us
- 17 9) Researcher: Lovely
- 18 10) Interviewee: But I think it's still a useful stimulus for a piece of academic English writing xxx and it
- 19 should still reveal... how do they put vocabulary and syntax available to them xxx and those who are
- 20 struggling
- 21 11) Researcher: Ok. That's for the first task. How about the second task where students are asked to
- 22 write an argumentative essay, do you think this is also related to the type of tasks students do at
- 23 university?
- 24 12) Interviewee: Yes, I do xxx em. I think what we probably xxx expect people to do with this kind of
- 25 task is to give xxx evidence for the points that they are presenting for their argument.
- 26 13) Researcher: Em
- 27 14) Interviewee: So you would expect people to make reference to the literature and so on xxx but again
- 28 even though the task is a little bit artificial because it's just asking for their opinion about a particular
- 29 idea
- 30 15) Researcher: Yep
- 31 16) Interviewee: It's likely writing task one it's still gives the context in which people can reveal their
- 32 abilities to write argumentative texts and that's certainly very relevant so yeah
- 33 17) Researcher: Lovely
- 34 18) Interviewee: I think the tasks are appropriate
- 35 19) Researcher: As a lecturer in this school you have probably met so many international students, can I
- 36 ask you what is your expectation of international students enrolled in your course since they have
- 37 completed this test and met the minimum requirements?
- 38 20) Interviewee: Yeah. The expectations in terms of standards of writing and in terms of standards of our
- 39 argument are the same as for students em xxx whose first language is English em but we certainly
- 40 expect to provide more support for international students in order to help them to reach to that
- 41 standard
- 42 21) Researcher: Good

## Lecturers' Perception of the IELTS Test Interview

// Interruptions

... Inaudible

xxx Pause

xxxxxx Long pause

- 43 22) Interviewee: And that's reflected in the work load model that governs what's staff do and there is  
44 more time available to supervise international students.
- 45 23) Researcher: Em
- 46 24) Interviewee: Technically students whose first language isn't English and it's also reflected in the  
47 kind of support services in the university xxx and I think probably I would expect xxx a certain  
48 amount of flexibility in the very early stages of the course so xxx em I would be less fussy about the  
49 quality of English in the first assignment but by the time we get to the final xxx //
- 50 25) Researcher: // Submission
- 51 26) Interviewee: Submission yeah, then that seizes to be an issue and standards have to be the same
- 52 27) Researcher: Oh, Ok that's great. Having in mind that students receive support at university and they  
53 might actually do a self-study at home, why do you think some students fail to cope with the  
54 demands of the university despite achieving the minimum requirements and support from lecturers?
- 55 28) Interviewee: I think there are so many different issues involved xxx one is that the language of xxx  
56 social science research is very technical I mean terms have very particular meanings which are not  
57 always very identical to the everyday language use of the word xxx reliability for example
- 58 29) Researcher: Ok yeah
- 59 30) Interviewee: And if students are able to write it at IELTS 6.5 level that doesn't necessarily mean that  
60 they have the flexibility to deal with that technical language
- 61 31) Researcher: Ok
- 62 32) Interviewee: At least that's going to be a difficulty. And that's a difficulty for English as a first  
63 language speakers
- 64 33) Researcher: Oh, is it?
- 65 34) Interviewee: Trying to understand what exactly the difference is between epistemology and ontology.  
66 Those are difficult issues em so xxx trying to understand those when English is your second  
67 language I think is an additional layer of complexity.
- 68 35) Researcher: Definitely
- 69 36) Interviewee: I think too there are issues about the way language is used and what's sort of  
70 acceptable. There is so much emphasis in the UK in higher degree courses in critical engagement and  
71 I think xxxxxx you probably know this more than I do but I think some international students feel  
72 uncomfortable about being critical of leading sources xxx being critical of what's said in lectures and  
73 so it's not just about the language but it's about the purpose to which this language is being put.
- 74 37) Researcher: Exactly and sometimes they have to follow a certain style of writing and are forced to  
75 write academically.
- 76 38) Interviewee: Yeah
- 77 39) Researcher: Now, When it comes to developing academic writing skills, what would you recommend  
78 teachers focus on with L2 writers?
- 79 40) Interviewee: I think there are different kinds of support. One one is to give a kind of writing frame  
80 which sits well in this piece you need to start by saying that and then that and then that so the  
81 students need a structure to work with. Another is to give detailed feedback on texts em and that's  
82 probably particularly in the early stages and em now I am quite happy to sort of work with a piece of  
83 draft material from a student and reword things and make suggestions about how it could be written  
84 more convincingly
- 85 41) Researcher: At the beginning

## Lecturers' Perception of the IELTS Test Interview

// Interruptions

... Inaudible

xxx Pause

xxxxxx Long pause

- 86 42) Interviewee: You actually expect people to become more//  
87 43) Researcher: //Advanced  
88 44) Interviewee: Advanced but then a different kind of support because in a way those are forms of  
89 support that are intended to help students to improve their English so they are free-standing writers  
90 of high quality academic English but another kind of support is to say well you can get help you can  
91 em you can employ xxx an editor to revise the English once you've got the argument right.  
92 45) Researcher: Right  
93 46) Interviewee: So it's helping people to find ways of getting round the problem if there is still a  
94 problem and I like to put more emphasis on support xxx that would enable people to acquire the skill  
95 47) Researcher: Ok  
96 48) Interviewee: But it's important to also to be able to get round the problem if there is still difficulties  
97 49) Researcher: I agree with you actually what you are saying is different from what the test claims.  
98 50) Interviewee: Oh  
99 51) Researcher: According to the IELTS handbook the test claims that students should be ready to start  
100 their academic courses at university. Do you they are ready?  
101 52) Interviewee: I think it's a very difficult question because xxxxxx I think they are as ready as they can  
102 be before they engage in the course itself. em There is a sense in which academic writing is a sort of  
103 general construct xxx and you can be trained to do that but as I was saying earlier some of the  
104 difficulties come from the kinds of concepts that are involved the kinds of language that is used to  
105 describe those concepts. And those are challenging for any speaker whether English is a first or  
106 second language and so to imagine that an international student is going to come into an NEE session  
107 in the philosophy of educational research and find that language perfectly straight forward is  
108 unrealistic  
109 53) Researcher: That's right  
110 54) Interviewee: That a struggle. So, I would see IELTS as providing the sort of grand level to say yes  
111 you are able to begin this journey but it is a journey and it will take time to become a competent and  
112 write in education. It would be a slightly different journey for somebody in Physics.  
113 55) Researcher: Absolutely. So you think it's preparing students for general English and not for  
114 academic English.  
115 56) Interviewee: I think it's preparing people for a general approach to academic writing but then that  
116 has to have xxx the context of the subject .... added to it .. all sorts of things come to mind xxx and I  
117 think in xxx subjects like Education we draw on literature across a wide range of disciplines and the  
118 language in all of these disciplines is slightly different and the technical usage is slightly different  
119 too. So, what a sociologist means by power is not necessarily what a xxx//  
120 57) Researcher: //An applied linguist  
121 58) Interviewee: Applied linguist might mean by power and that adds another layer of difficulty for  
122 education studies students. Physics is a more discrete discipline em once you are a physicist the  
123 language of Physics is fairly contained and the language of Education is the language of Psychology  
124 the language of Sociology the language of Philosophy the language xxx of Education and each of  
125 those makes slightly different demands different kinds of pitfalls  
126 59) Researcher: I agree with you IELTS is now measuring general academic English as those who are  
127 developing it and doing validation studies on it found that there are no boundaries between these  
128 disciplines. Anyway, as you know advanced writing especially at university level requires students to

## Lecturers' Perception of the IELTS Test Interview

// Interruptions

... Inaudible

xxx Pause

xxxxxx Long pause

- 129 produce complex structures and formal language, the assessment process on the part of the teachers  
130 then perhaps becomes equally demanding, what do you think the criteria should be at university I  
131 mean compared to this test, do you think there are some sort of similarities and differences between  
132 IELTS tasks and university tasks?
- 133 60) Interviewee: xxxxxxxx It's a really difficult question. I think that it's unrealistic to expect an  
134 IELTS test to cover the same criteria as university assignments because if you were to do that you  
135 would have to be based on educational content in the preparation for that test. So in fact students  
136 would have to do the course before they come to do the course
- 137 61) Researcher: That's impossible
- 138 62) Interviewee: And xxx where that takes me I think really is to say that universities have recognise that  
139 an IELTS score doesn't guarantee that an individual is xxxxxx sort of already an expert in the field  
140 they are going to study. If they've done a Masters in the Education overseas and the nature of that  
141 stresses different kinds of abilities and different kinds of reasoning and different kinds of writing.  
142 Then, just having done an IELTS test to say yes your academic English is good
- 143 63) Researcher: Ok
- 144 64) Interviewee: Doesn't mean that the student is starting from the same position on a PhD programme  
145 as a UK student who has done a Masters that already has tuned them up to the kind of writing  
146 expected in the UK Masters level. Now, I think the difference is between xxx say em a Chinese  
147 Masters and English Masters may not be enormous but those differences will make an impact on  
148 how the student is able to progress in their early stages of their course in a UK university and I don't  
149 think the IELTS system can do that bridging. The IELTS system can make sure that they are able to  
150 do the kinds of argument that they used to do but do them in English but it doesn't necessarily mean  
151 that they can do the sort of arguments that they need to do in an English course.
- 152 65) Researcher: Exactly, so do you think that this would threaten the predictive validity of the test. I  
153 mean from these two tasks they can generalise that students would be able to do longer ones.
- 154 66) Interviewee: Mind you people do say it's easier to write a long piece than a short piece
- 155 67) Researcher: Maybe
- 156 68) Interviewee: I think the main jest of my argument is that we can't expect scores on an IELTS test to  
157 guarantee that the student is in exactly the same place to begin their postgraduate study as a student  
158 who has done their education in the UK because in addition to having a Masters degree and the  
159 ability to write in English the international student has to acquire the ability to xxx write in the styles  
160 and within the culture of UK higher Education. And those two separate things I mean the overseas  
161 Masters and the IELTS don't put you quite in the same place as someone who has come through the  
162 UK system. I think that gives the international student some advantages because when you compare  
163 and contrast the different experience with the new experience you can create new understanding  
164 which the UK student struggles to find because you can start to see what we used to do like that now  
165 we are doing it like this. But at the same time it makes the task quite challenging because there is that  
166 cultural shift
- 167 69) Researcher: I see
- 168 70) Interviewee: To do
- 169 71) Researcher: So do you think they are claiming too much from this simple task?

## Lecturers' Perception of the IELTS Test Interview

// Interruptions

... Inaudible

xxx Pause

xxxxxx Long pause

- 170 72) Interviewee: I think if IELTS says that this is a test of general or academic writing it's probably a fair  
171 claim. What I think universities have to be careful about is assuming that that turns an international  
172 student into someone who is in the position of exactly the same as a home student.//  
173 73) Researcher: //So do you think that the problem is with universities who may be misinterpreting the  
174 overall IELTS scores  
175 74) Interviewee: I think they expect too much from it. Perhaps not engaging enough with xxx that  
176 cultural shift or supporting that cultural shift  
177 75) Researcher: I do agree with in this regard because universities have different minimum scores for  
178 IELTS. Now, when they have problems with underachieving students in English they raise the scores  
179 for next year. Do you think that raising the IELTS score would sift the good students?  
180 76) Interviewee: You are describing exactly the same point that I wanted to make. A conclusion of my  
181 argument is that raising the IELTS threshold wouldn't help because you still got that cultural shift to  
182 support and it might help marginally but they are missing the point in raising it from 6.5 to 7.  
183 77) Researcher: This statement brings us to the end of this interview. It's pleasure having you.  
184 78) Interviewee: It's nice to be involved  
185 79) Researcher: Thank you very for your participation  
186 80) Interviewee: You are welcome  
187

## **Key Findings (according to data collection tools)**

### **Phase I**

#### **Key findings for the students**

##### **(a) Students' Pre-Writing Test Questionnaires**

- 1- Whether the students had similar or different purposes for taking the tests, it seems that the majority were more likely to be driven by an institutional choice rather than a personal one.
- 2- The value of the test was seen by the students to be as a means of access more than as a measure of proficiency.
- 3- It seems that the more demanding the course, either postgraduate or undergraduate, in terms of communicative competence the higher the IELTS scores should be. These students were applying to different courses in different universities and countries, and there may be some sort of an agreement over the lowest scores at least among this group of students.
- 4- There was fairly high level of awareness of test content among these students. Ultimately, this high level of awareness may have a positive impact on their performance in that it may help them to successfully apply certain test-taking strategies e.g. time management.
- 5- Students' responses indicated that they would focus on different things at the beginning of the writing test.
- 6- While writing, most students indicated that they would pay attention to task achievement.
- 7- The majority of the students reported that they would consider planning and revision strategies to be the most important strategies in the writing test.
- 8- Time limitations and unfamiliar topics were issues of concern to these students in the writing test in IELTS. These aspects may, as a result, hinder the students from showing their best performance in the test in a way or another and therefore we may risk having an insufficient evidence of what the students are able to accomplish in the target situation under such restrictions.

## **(b) Students' Post-Writing Test Questionnaires**

- 1- Students had mixed feelings about the suggested duration of the writing test and the first task in particular. Yet, it seems that almost all of them would agree that the allotted time for the second task was enough.
- 2- Students valued the importance of instructions but needed more guidance in order to focus their attention on the important aspects of the task which in turn may help them to allocate the appropriate time for each part, an issue which was always of concern to them.
- 3- Despite the time limitations and the nature of the test, most of these students have made use of brainstorming to organise their ideas. This could in turn keep them focused on the topic and help them to write relevant information.
- 4- Most of the students indicated that the first task presented a challenge to them because there was too much information in the rubric which caused confusion, and it was difficult to establish a connection between the table and the chart. Yet, some of them found the words of the questions of both tasks to be free of any problems causing misunderstanding or confusion.
- 5- There was a variation in the students' responses about the frequency of going back to correct grammar mistakes, style and organisation during and after the writing process. This could be attributed to the kind of task, its requirements, and the time limitations associated with it.
- 6- The majority of the students reported that they were less familiar with the content of the writing tasks especially the first task. It seems that familiarity with the content meant different things to these students. Some have considered it to be familiarity with the task or the question and how to answer it by following a specific genre. On the other hand, other students have seen familiarity as being able to know the content of the task by checking the given information in the task against their prior knowledge about the subject or the content of the task.

### **(c) Students' Observations**

- 1- The researcher found that students spent different times while responding to test tasks.
- 2- The students occasionally referred to the prompt of the tasks to make sure that they covered every aspect in the question.
- 3- The students were scanning for the keywords in the questions and as soon as they identified them and understood the requirements of the tasks they started writing.
- 4- It seems that the delay time while responding to test tasks may have affected the use of certain test taking strategies.
- 5- Due to time restrictions, some students were seen to be planning and revising as they were writing.
- 6- The students have approached the tasks differently and very often their choice of test taking strategies was deliberate.

### **(d) Students' Interviews**

- 1- Most students reported that they were unable to make drafts especially in the first task and some of them could not even complete the task because of time restrictions.
- 2- Students found that the current guidelines especially in the first task were not helpful as they left a number of things implicit such as the relationship between the graph and the table.
- 3- Students stressed the importance of brainstorming in enhancing the quality of their writing. Yet, some of them were not able to apply it because of time restrictions or task type. Most of these students justified that the second task needed more

- brainstorming than the first one. Other students, however, overcame these problems and brainstormed in both tasks. This clearly showed that where one strategy would seem to be helpful to somebody, it could be, however, less useful and time consuming for somebody else due to certain restrictions or requirements.
- 4- Students have set priorities in fulfilling the criteria of the writing test and they were keen to first meet those criteria that would seem to be less controversial to examiners.
  - 5- Students were conscious in choosing the writing strategies. That is to say the cognitive processing activities addressing and linking various stages during the production of writing did not happen by chance.
  - 6- The first task was seen to present more challenge to most of these participants due to either time limitations, stress, difficulty in understanding the graph and the table, too much information in the graph and the table versus little information in the rubric, or missing information. On the other hand, the second task was also demanding to the rest of the students in terms of bearing more marks and words, doing two things at a time, and bringing outside information.
  - 7- Students were happy with the current order of tasks.
  - 8- Some students indicated that the more detailed the question of the task the more ideas they will have.
  - 9- Students have shown that writing relevant information would require a careful analysis of the question to understand the demands of the task.
  - 10- Some candidates have shown that some textual features in the prompts have aided them to meet two important criteria in marking academic writing tasks in IELTS namely; Task Achievement (in Task 1) and Task Response (in Task 2). They had to maintain a balance between these two criteria and among the rest of the assessment criteria in the writing test if successful fulfilment of task requirements was sought out.
  - 11- Some students have demonstrated that the test perhaps focused not only on language ability needed for survival in the target language use (TLU) context but also on topical knowledge manifested in their ability to reflect on their own experiences in the second task. This, issue raised about topical knowledge, may

explain why some students would consider the second task easier than the first one and more relevant to real life contexts.

**12-** Some students suggested that:

**A-** If writing test tasks were to reflect real life situations in measuring communicative ability, they should be designed by test developers to account for every possible response which would be accepted in target contexts and not just the one which existed in the mind of the examiner i.e. the model answer. Therefore, test developers should emphasise their own desired model examples which would be highly appreciated by IELTS examiners and to give less attention to those examples written by candidates. In doing so, the test would become more transparent to them through test preparation materials and the expected response could be highly predictable.

**B-** Some students may successfully get the required overall score but they may sometimes fail to meet the minimum requirements of one or more sub-scores. Therefore, it was suggested that in this case students should repeat just the modules that they failed in and not the whole test. The reason for such concern was because it was thought that test scores (either the overall score or the sub-score) may fluctuate after repeating the test.

**C-** There should be specific modules for each discipline as they sometimes faced difficult questions which were not related to their field of study.

## **Key Findings for the Marker's Questionnaire**

- 1- Expert markers don't need to have the scoring guide during the marking process because the marker has already internalised it.
- 2- Different markers may interpret the rating scales differently.
- 3- Training courses were found to be essential for gaining consistency among markers.
- 4- There were some attitudes e.g. leniency or severity that may be challenging to overcome for markers despite the fact that they had training sessions.
- 5- Some candidates may not be allocated correctly to their corresponding band score or level of performance because of leniency or severity.
- 6- Single marking could be adequate because IELTS is marked through an analytic scale.
- 7- The IELTS marker was not accurate when listing some of the criteria for scoring both tasks. That could be because she was scoring other tests which are similar to IELTS. However, the quality of scoring was not affected because she already internalised the scoring process of each task.
- 7- Markers need to be aware of several factors that may not be a threat to scoring validity but to the reliability of test scores namely; lack of expertise, misinterpretation of scoring criteria and extreme leniency or severity in scoring.

## **Phase II**

### **Key findings for the students**

#### **(a) Students' Questionnaires**

- 1- Most of the students have found the university tasks to be different from IELTS tasks in a number of different areas such as level of difficulty, variety of topics, styles, discourse mode, length, and language.
- 2- The students indicated that university tasks were more academic (vocabulary), longer in terms of time and word count, related to one's area of study (topics), and university writing demanded doing research which would involve reading references, and conducting surveys to collect data and materials for the study.
- 3- The students found IELTS to be more general (topics & vocabulary), much easier, very structured, and shorter in terms of time and word count.
- 4- The students found both tasks to be similar in several areas such as language i.e. academic English, discourse mode e.g. descriptive and argumentative, style e.g. objective writing, structure, cognitive demands e.g. analyse and write, organisation, grammar and vocabulary.
- 5- Most students indicated that IELTS writing test was somehow suitable but may not be enough alone if students wanted to develop those skills (e.g. referencing, critical thinking, managing long papers or assignments...etc.) needed for academic writing.

#### **(b) Students' Interviews**

- 1- The students noted that IELTS tasks and university tasks shared some textual features but with different expectations which may have caused the students to think that there was a weak relationship between IELTS tasks and university tasks.
- 2- The students found that they were lacking essential academic (research) skills needed for success at university and that the IELTS test offered them limited support in this regard.
- 3- The students have found that IELTS tasks shared some common features of target language tasks. Nonetheless, they were doubtful of the representativeness of

IELTS tasks as they may not exactly mirror the content and skills needed to undertake university tasks.

- 4- The students felt that more improvement was needed in areas such as format, content or scoring method.

### **Key Findings for Lecturers' Interviews**

- 1- Some lecturers expressed their concerns over the issue of readiness that is claimed by IELTS and therefore were doubtful that students would be able to cope with the demands of the university.
- 2- IELTS alone is not enough to prepare students for university and therefore universities should offer more support during the academic programmes. This is because the demands of writing an assignment at university are different from those in the IELTS writing test.
- 3- IELTS gives a rough measure of what a candidate is able to do and therefore it may be difficult to generalise on the students' performance in these tasks.
- 4- University courses require high thinking and critical writing skills while IELTS would equip students with the technical elements (e.g coherence and cohesion...etc.).
- 5- IELTS is best suited for preparing undergraduate students for year one because this is the kind of activities they will be doing. In this regard, the test could be misused.
- 6- Most lecturers found that there was low level of correspondence between IELTS tasks and university tasks.
- 7- Most lecturers were not happy about raising the IELTS scores because it wouldn't ensure that only good students would be accepted into university as the difference in performance would be so little and may not be noticed.
- 8- Some lecturers noticed that there was low correlation between IELTS scores and what a candidate can achieve.
- 9- Most lecturers expected students to be at a level in English where they can survive at university while some of them were disappointed because the test did not match

their expectations and thus students needed to have more support in order to cope with the university demands.

**10-**Most lecturers reported that they would be tolerant when it comes to offering advice on a piece of writing regarding structure, organisation, being critical, developing arguments...etc. but they would leave linguistic errors for a proof-reader.

## Timeframe of the Research

<b>Timeframe of Research (Ahmad Alsagoafi)</b>										
	Nov 2009- January 2010	Feb 2010- March 2010	April 2010- June 2010	June 2010-July 2010	Aug 2010- Oct 2010	Nov 2010- Feb 2011	March 2011- May 2011	June 2011-Sep 2011	Oct 2011- Feb 2012	March 2012-Sep 2012
<b>Research Methodology</b>	Develop research proposal and design questions			Review research methodology						
<b>Literature Review</b>	Literature review into academic writing in IELTS, testing and assessing academic writing in EFL/ESL contexts with focus on issues to do with validity, impact, washback, authenticity, meaning-making, task design and scoring. Theories related to teaching and testing second as well as foreign language performance.									
<b>Sampling</b>	Initial contact with the British Council	Locate the participants for the pilot study	Conducting questionnaires, interviews and observations for the pilot study							
<b>Data Collection</b>		Doing questionnaires and interviews of pilot participants	1 <sup>st</sup> Phase data collection (Q,I&O)KSA	2 <sup>nd</sup> Phase data collection (Q,I&O) UK						
<b>Data Analysis</b>			Data analysis of the pilot study	Data analysis for 1 <sup>st</sup> Phase	Data analysis for 2 <sup>nd</sup> Phase					
<b>Instruments</b>		Designing questionnaire, interview guidelines and observation schedules	Review of the questionnaires, interview guidelines and observation schedules							
<b>Writing Up</b>			Methodology & literature review (1 <sup>st</sup> draft)	Draft review (after feedback)	Data analysis for 1 <sup>st</sup> Phase	Data analysis for 2 <sup>nd</sup> Phase & Methodology (2 <sup>nd</sup> draft)	Literature review (2 <sup>nd</sup> draft) & conclusions	Finalised work and submission		



**GRADUATE SCHOOL OF EDUCATION**

**CONSENT FORM**

I have been fully informed about the aims and purposes of the project.

I understand that:

there is no compulsion for me to participate in this research project and, if I do choose to participate, I may at any stage withdraw my participation

I have the right to refuse permission for the publication of any information about me

any information which I give will be used solely for the purposes of this research project, which may include publications

If applicable, the information, which I give, may be shared between any of the other researcher(s) participating in this project in an anonymised form

all information I give will be treated as confidential  
the researcher(s) will make every effort to preserve my anonymity

.....

(Signature of participant )

.....

(Date)

.....

(Printed name of participant)

One copy of this form will be kept by the participant; a second copy will be kept by the researcher(s)

Contact phone number of researcher(s):.....

If you have any concerns about the project that you would like to discuss, please contact:

..... OR .....

Data Protection Act: The University of Exeter is a data collector and is registered with the Office of the Data Protection Commissioner as required to do under the Data Protection Act 1998. The information you provide will be used for research purposes and will be processed in accordance with the University's registration and current data protection legislation. Data will be confidential to the researcher(s) and will not be disclosed to any unauthorised third parties without further agreement by the participant. Reports based on the data will be in anonymised form.

## Appendix II

### IELTS TASKS & SCORING GUIDES

#### Sample of IELTS Writing Task 1

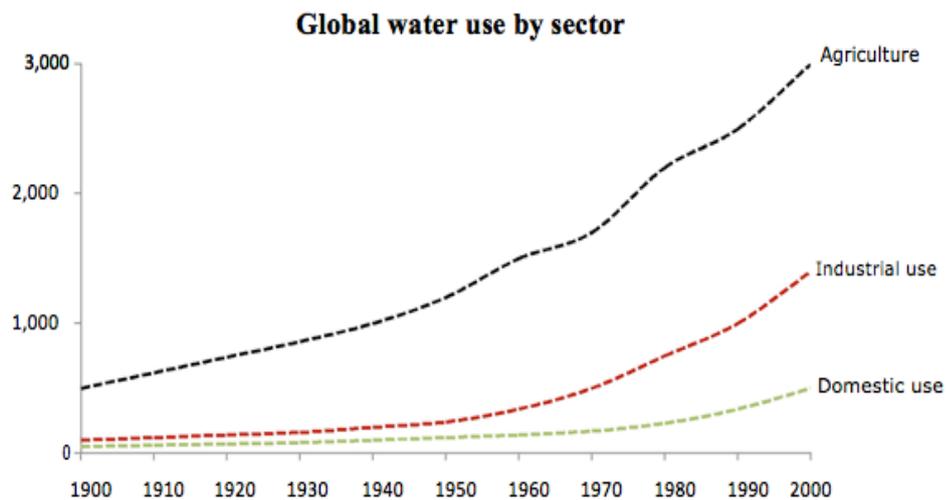
##### WRITING TASK 1

You should spend about 20 minutes on this task.

*The graph and table below give information about water use worldwide and water consumption in two different countries.*

*Summarise the information by selecting and reporting the main features, and make comparisons where relevant.*

Write at least 150 words.



**Water consumption in Brazil and Congo in 2000**

Country	Population	Irrigated land	Water consumption per person
Brazil	176 million	26,500 km <sup>2</sup>	359 m <sup>3</sup>
Democratic Republic of Congo	5.2 million	100 km <sup>2</sup>	8 m <sup>3</sup>

## Model answer for IELTS Writing Task 1

### **Model answer**

This model has been prepared by an examiner as an example of a very good answer. However, please note that this is just one example out of many possible approaches.

The graph shows how the amount of water used worldwide changed between 1900 and 2000.

Throughout the century, the largest quantity of water was used for agricultural Purposes, and this increased dramatically from about 500 km<sup>3</sup> to around 3,000 km<sup>3</sup> in the year 2000. Water used in the industrial and domestic sectors also increased, but consumption was minimal until mid-century. From 1950 onwards, industrial use grew steadily to just over 1,000 km<sup>3</sup>, while domestic use rose more slowly to only 300 km<sup>3</sup>, both far below the levels of consumption by agriculture.

The table illustrates the differences in agriculture consumption in some areas of the world by contrasting the amount of irrigated land in Brazil (26,500 km<sup>2</sup>) with that in the D.R.C. (100 km<sup>2</sup>). This means that a huge amount of water is used in agriculture in Brazil, and this is reflected in the figures for water consumption per person: 359 m<sup>3</sup> compared with only 8 m<sup>3</sup> in the Congo. With a population of 176 million, the figures for Brazil indicate how high agriculture water consumption can be in some countries.

Retrieved from IELTS Exam Preparation 2007 <http://www.ielts-exam.net>

## Sample of IELTS Writing Task 2

### WRITING TASK 2

You should spend about 40 minutes on this task.

Write about the following topic:

*Some people prefer to spend their lives doing the same things and avoiding change. Others, however, think that change is always a good thing.*

*Discuss both these views and give your own opinion.*

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

## Model answer for IELTS Writing Task 2

### **Model answer**

This model has been prepared by an examiner as an example of a very good answer. However, please note that this is just one example out of many possible approaches.

Over the last half century the pace of change in the life of human beings has increased beyond our wildest expectations. This has been driven by technological and scientific breakthroughs that are changing the whole way we view the world on an almost daily basis. This means that change is not always a personal option, but an inescapable fact of life, and we need to constantly adapt to keep pace with it.

Those people who believe they have achieved some security by doing the same, familiar things are living in denial. Even when people believe they are resisting change themselves, they cannot stop the world around them from changing. Sooner or later they will find that the familiar jobs no longer exist, or that the 'safe' patterns of behaviour are no longer appropriate.

However, reaching the conclusion that change is inevitable is not the same as assuming that 'change is always for the better'. Unfortunately, it is not always the case that new things are promoted because they have good impacts for the majority of people. A lot of innovations are made with the aim of making money for a few. This is because it is the rich and powerful people in our society who are able to impose changes (such as in working conditions or property developments) that are in their own interests.

In conclusion, I would say that change can be stimulating and energising for individuals when they pursue it themselves, but that all change, including that which is imposed on people, does not necessarily have good outcomes.

IELTS Exam Preparation (2007) Retrieved from: <http://www.ielts-exam.net>

## Scoring Criteria for Task One

<b>5</b>	<ul style="list-style-type: none"> <li>■ generally addresses the task; the format may be inappropriate in places</li> <li>■ (Academic) recounts detail mechanically with no clear overview; there may be no data to support the description</li> <li>■ (General Training) may present a purpose for the letter that is unclear at times; the tone may be variable and sometimes inappropriate</li> <li>■ presents, but inadequately covers, key features / bullet points; there may be a tendency to focus on details</li> </ul>	<ul style="list-style-type: none"> <li>■ presents information with some organisation but there may be a lack of overall progression</li> <li>■ makes inadequate, inaccurate or over-use of cohesive devices</li> <li>■ may be repetitive because of lack of referencing and substitution</li> </ul>	<ul style="list-style-type: none"> <li>■ uses a limited range of vocabulary, but this is minimally adequate for the task</li> <li>■ may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only a limited range of structures</li> <li>■ attempts complex sentences but these tend to be less accurate than simple sentences</li> <li>■ may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader</li> </ul>
<b>4</b>	<ul style="list-style-type: none"> <li>■ attempts to address the task but does not cover all key features / bullet points; the format may be inappropriate</li> <li>■ (General Training) fails to clearly explain the purpose of the letter; the tone may be inappropriate</li> <li>■ may confuse key features / bullet points with detail; parts may be unclear, irrelevant, repetitive or inaccurate</li> </ul>	<ul style="list-style-type: none"> <li>■ presents information and ideas but these are not arranged coherently and there is no clear progression in the response</li> <li>■ uses some basic cohesive devices but these may be inaccurate or repetitive</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task</li> <li>■ has limited control of word formation and/or spelling</li> <li>■ errors may cause strain for the reader</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only a very limited range of structures with only rare use of subordinate clauses</li> <li>■ some structures are accurate but errors predominate; and punctuation is often faulty</li> </ul>
<b>3</b>	<ul style="list-style-type: none"> <li>■ fails to address the task, which may have been completely misunderstood</li> <li>■ presents limited ideas which may be largely irrelevant/repetitive</li> </ul>	<ul style="list-style-type: none"> <li>■ does not organise ideas logically</li> <li>■ may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only a very limited range of words and expressions with very limited control of word formation and/or spelling</li> <li>■ errors may severely distort the message</li> </ul>	<ul style="list-style-type: none"> <li>■ attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning</li> </ul>
<b>2</b>	<ul style="list-style-type: none"> <li>■ answer is barely related to the task</li> </ul>	<ul style="list-style-type: none"> <li>■ has very little control of organisational features</li> </ul>	<ul style="list-style-type: none"> <li>■ uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling</li> </ul>	<ul style="list-style-type: none"> <li>■ cannot use sentence forms except in memorised phrases</li> </ul>
<b>1</b>	<ul style="list-style-type: none"> <li>■ answer is completely unrelated to the task</li> </ul>	<ul style="list-style-type: none"> <li>■ fails to communicate any message</li> </ul>	<ul style="list-style-type: none"> <li>■ can only use a few isolated words</li> </ul>	<ul style="list-style-type: none"> <li>■ cannot use sentence forms at all</li> </ul>
<b>0</b>	<ul style="list-style-type: none"> <li>■ does not attend</li> <li>■ does not attempt the task in any way</li> <li>■ writes a totally memorised response</li> </ul>			

Band	Task Achievement	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
9	<ul style="list-style-type: none"> <li>fully satisfies all the requirements of the task</li> <li>clearly presents a fully developed response</li> </ul>	<ul style="list-style-type: none"> <li>uses cohesion in such a way that it attracts no attention</li> <li>skillfully manages paragraphing</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'</li> </ul>
8	<ul style="list-style-type: none"> <li>covers all requirements of the task sufficiently</li> <li>presents, highlights and illustrates key features / bullet points clearly and appropriately</li> </ul>	<ul style="list-style-type: none"> <li>sequences information and ideas logically</li> <li>manages all aspects of cohesion well</li> <li>uses paragraphing sufficiently and appropriately</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of vocabulary fluently and flexibly to convey precise meanings</li> <li>skillfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation</li> <li>produces rare errors in spelling and/or word formation</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of structures</li> <li>the majority of sentences are error-free</li> <li>makes only very occasional errors or inaccuracies</li> </ul>
7	<ul style="list-style-type: none"> <li>covers the requirements of the task</li> <li>(Academic) presents a clear overview of main trends, differences or stages</li> <li>(General Training) presents a clear purpose, with the tone consistent and appropriate</li> <li>clearly presents and highlights key features / bullet points but could be more fully extended</li> </ul>	<ul style="list-style-type: none"> <li>logically organises information and ideas; there is clear progression throughout</li> <li>uses a range of cohesive devices appropriately although there may be some under-/over-use</li> </ul>	<ul style="list-style-type: none"> <li>uses a sufficient range of vocabulary to allow some flexibility and precision</li> <li>uses less common lexical items with some awareness of style and collocation</li> <li>may produce occasional errors in word choice, spelling and/or word formation</li> </ul>	<ul style="list-style-type: none"> <li>uses a variety of complex structures</li> <li>produces frequent error-free sentences</li> <li>has good control of grammar and punctuation but may make a few errors</li> </ul>
6	<ul style="list-style-type: none"> <li>addresses the requirements of the task</li> <li>(Academic) presents an overview with information appropriately selected</li> <li>(General Training) presents a purpose that is generally clear; there may be inconsistencies in tone</li> <li>presents and adequately highlights key features / bullet points but details may be irrelevant, inappropriate or inaccurate</li> </ul>	<ul style="list-style-type: none"> <li>arranges information and ideas coherently and there is a clear overall progression</li> <li>uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical</li> <li>may not always use referencing clearly or appropriately</li> </ul>	<ul style="list-style-type: none"> <li>uses an adequate range of vocabulary for the task</li> <li>attempts to use less common vocabulary but with some inaccuracy</li> <li>makes some errors in spelling and/or word formation, but they do not impede communication</li> </ul>	<ul style="list-style-type: none"> <li>uses a mix of simple and complex sentence forms</li> <li>makes some errors in grammar and punctuation but they rarely reduce communication</li> </ul>

IELTS (2010a) Score Processing and Reporting. Retrieved from [http://www.ielts.org/researchers/score\\_processing\\_and\\_reporting.aspx](http://www.ielts.org/researchers/score_processing_and_reporting.aspx).

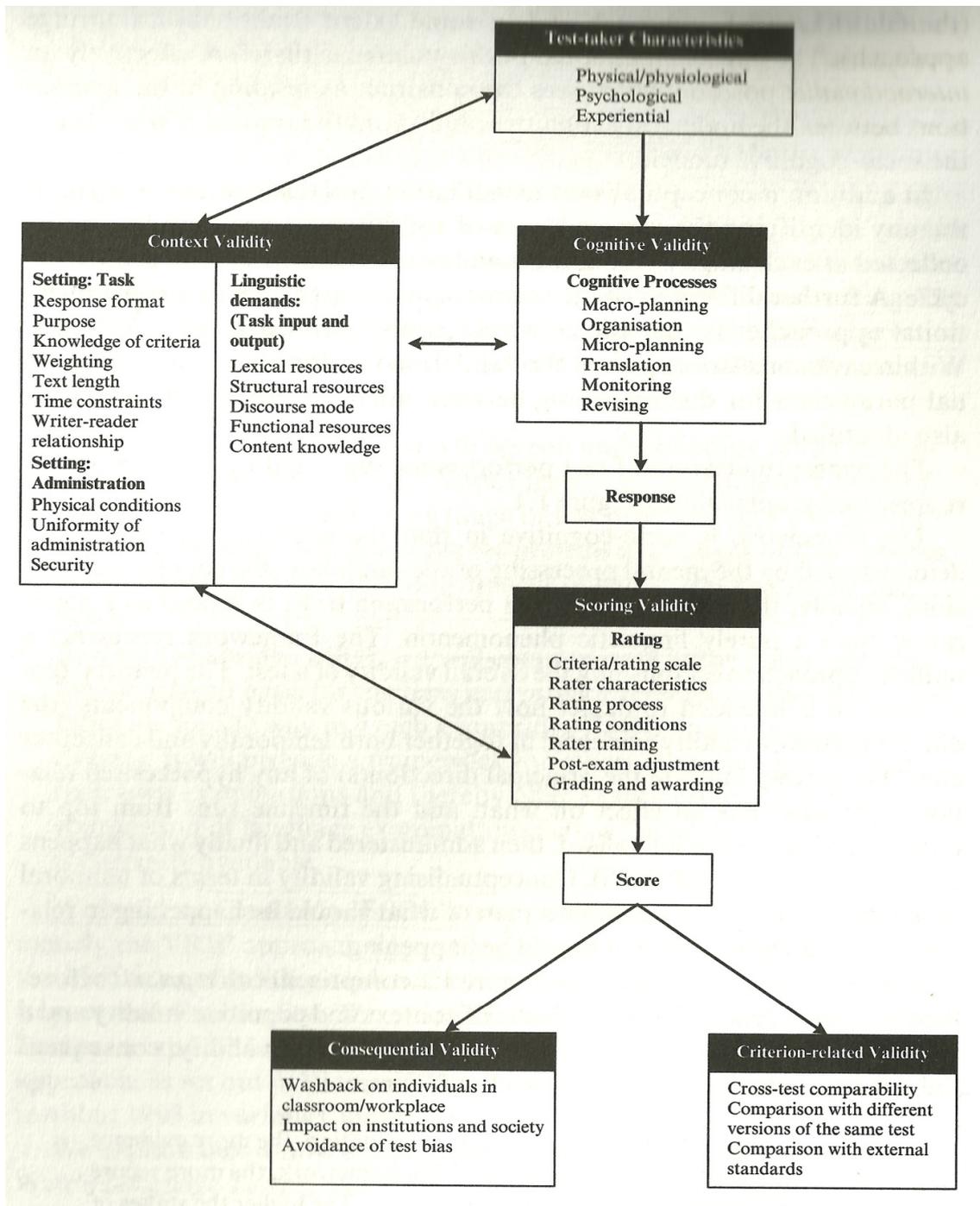
## Scoring Criteria for Task Two

<p>conclusions drawn</p> <ul style="list-style-type: none"> <li>■ presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail</li> </ul>	<p>use of cohesive devices</p> <ul style="list-style-type: none"> <li>■ may be repetitive because of lack of referencing and substitution</li> <li>■ may not write in paragraphs, or paragraphing may be inadequate</li> </ul>	<p>spelling and/or word formation that may cause some difficulty for the reader</p>	<ul style="list-style-type: none"> <li>■ may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader</li> </ul>
<p>4</p> <ul style="list-style-type: none"> <li>■ responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate</li> <li>■ presents a position but this is unclear</li> <li>■ presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported</li> </ul>	<ul style="list-style-type: none"> <li>■ presents information and ideas but these are not arranged coherently and there is no clear progression in the response</li> <li>■ uses some basic cohesive devices but these may be inaccurate or repetitive</li> <li>■ may not write in paragraphs or their use may be confusing</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task</li> <li>■ has limited control of word formation and/or spelling; errors may cause strain for the reader</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only a very limited range of structures with only rare use of subordinate clauses</li> <li>■ some structures are accurate but errors predominate, and punctuation is often faulty</li> </ul>
<p>3</p> <ul style="list-style-type: none"> <li>■ does not adequately address any part of the task</li> <li>■ does not express a clear position</li> <li>■ presents few ideas, which are largely undeveloped or irrelevant</li> </ul>	<ul style="list-style-type: none"> <li>■ does not organise ideas logically</li> <li>■ may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas</li> </ul>	<ul style="list-style-type: none"> <li>■ uses only a very limited range of words and expressions with very limited control of word formation and/or spelling</li> <li>■ errors may severely distort the message</li> </ul>	<ul style="list-style-type: none"> <li>■ attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning</li> </ul>
<p>2</p> <ul style="list-style-type: none"> <li>■ barely responds to the task</li> <li>■ does not express a position</li> <li>■ may attempt to present one or two ideas but there is no development</li> </ul>	<ul style="list-style-type: none"> <li>■ has very little control of organisational features</li> </ul>	<ul style="list-style-type: none"> <li>■ uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling</li> </ul>	<ul style="list-style-type: none"> <li>■ cannot use sentence forms except in memorised phrases</li> </ul>
<p>1</p> <ul style="list-style-type: none"> <li>■ answer is completely unrelated to the task</li> </ul>	<ul style="list-style-type: none"> <li>■ fails to communicate any message</li> </ul>	<ul style="list-style-type: none"> <li>■ can only use a few isolated words</li> </ul>	<ul style="list-style-type: none"> <li>■ cannot use sentence forms at all</li> </ul>
<p>0</p>	<ul style="list-style-type: none"> <li>■ does not attend</li> <li>■ does not attempt the task in any way</li> <li>■ writes a totally memorised response</li> </ul>		

Band	Task Response	Coherence and Cohesion	Lexical Resource	Grammatical Range and Accuracy
9	<ul style="list-style-type: none"> <li>fully addresses all parts of the task</li> <li>presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas</li> </ul>	<ul style="list-style-type: none"> <li>uses cohesion in such a way that it attracts no attention</li> <li>skillfully manages paragraphing</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'</li> </ul>
8	<ul style="list-style-type: none"> <li>sufficiently addresses all parts of the task</li> <li>presents a well-developed response to the question with relevant, extended and supported ideas</li> </ul>	<ul style="list-style-type: none"> <li>sequences information and ideas logically</li> <li>manages all aspects of cohesion well</li> <li>uses paragraphing sufficiently and appropriately</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of vocabulary fluently and flexibly to convey precise meanings</li> <li>skillfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation</li> <li>produces rare errors in spelling and/or word formation</li> </ul>	<ul style="list-style-type: none"> <li>uses a wide range of structures</li> <li>the majority of sentences are error-free</li> <li>makes only very occasional errors or inappropriacies</li> </ul>
7	<ul style="list-style-type: none"> <li>addresses all parts of the task</li> <li>presents a clear position throughout the response</li> <li>presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus</li> </ul>	<ul style="list-style-type: none"> <li>logically organises information and ideas, there is clear progression throughout</li> <li>uses a range of cohesive devices appropriately although there may be some under-/over-use</li> <li>presents a clear central topic within each paragraph</li> </ul>	<ul style="list-style-type: none"> <li>uses a sufficient range of vocabulary to allow some flexibility and precision</li> <li>uses less common lexical items with some awareness of style and collocation</li> <li>may produce occasional errors in word choice, spelling and/or word formation</li> </ul>	<ul style="list-style-type: none"> <li>uses a variety of complex structures</li> <li>produces frequent error-free sentences</li> <li>has good control of grammar and punctuation but may make a few errors</li> </ul>
6	<ul style="list-style-type: none"> <li>addresses all parts of the task although some parts may be more fully covered than others</li> <li>presents a relevant position although the conclusions may become unclear or repetitive</li> <li>presents relevant main ideas but some may be inadequately developed/unclear</li> </ul>	<ul style="list-style-type: none"> <li>arranges information and ideas coherently and there is a clear overall progression</li> <li>uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical</li> <li>may not always use referencing clearly or appropriately</li> <li>uses paragraphing, but not always logically</li> </ul>	<ul style="list-style-type: none"> <li>uses an adequate range of vocabulary for the task</li> <li>attempts to use less common vocabulary but with some inaccuracy</li> <li>makes some errors in spelling and/or word formation, but they do not impede communication</li> </ul>	<ul style="list-style-type: none"> <li>uses a mix of simple and complex sentence forms</li> <li>makes some errors in grammar and punctuation but they rarely reduce communication</li> </ul>
5	<ul style="list-style-type: none"> <li>addresses the task only partially, the format may be inappropriate in places</li> <li>expresses a position but the development is not always clear and there may be no</li> </ul>	<ul style="list-style-type: none"> <li>presents information with some organisation but there may be a lack of overall progression</li> <li>makes inadequate, inaccurate or over-</li> </ul>	<ul style="list-style-type: none"> <li>uses a limited range of vocabulary, but this is minimally adequate for the task</li> <li>may make noticeable errors in</li> </ul>	<ul style="list-style-type: none"> <li>uses only a limited range of structures</li> <li>attempts complex sentences but these tend to be less accurate than simple sentences</li> </ul>

IELTS (2010a) Score Processing and Reporting. Retrieved from [http://www.ielts.org/researchers/score\\_processing\\_and\\_reporting.aspx](http://www.ielts.org/researchers/score_processing_and_reporting.aspx).

**Appendix III**  
**Weir's Socio-cognitive Framework for Writing**



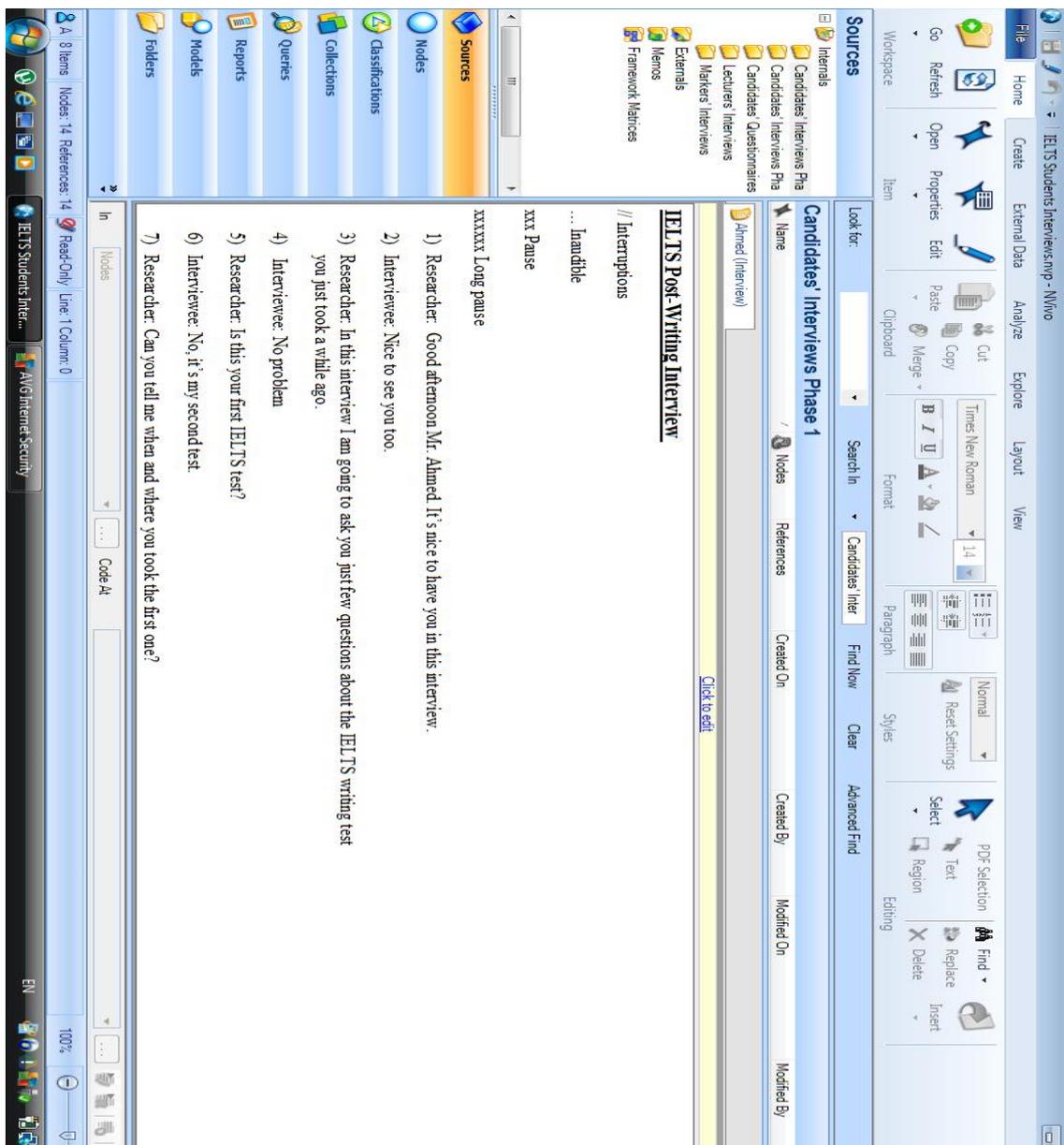
A framework for conceptualising writing test performance. Obtained from Shaw and Weir (2007) p.4.

## Appendix IV NVivo Samples

The screenshot displays the NVivo interface with the 'Nodes' pane open. The nodes list includes the following items:

- Writing relevant information
- Detailed questions
- Task difficulty and order
- Reading the question
- Paraphrasing the question
- Validity
- Suggestions
- Correcting mistakes
- Relating the topic of the task to what sb knows
- Preparation
- Comparing IELTS with TOEFL
- Time to make drafts
- Test Taking
- TOEFL
- Comparing the 1st task with the 2nd
- Strategies
- Exerting effort
- Guidelines
- Brainstorming
- Expectation of Uni Tasks
- IELTS different from Uni
- IELTS similar to Uni
- The relationship between IELTS writing tasks and university tasks
- The role of IELTS in preparing students for uni tasks
- Students' perspectives on the level of authenticity of IELTS writing tasks
- Suggestions Phase II

An example of the Nodes in NVivo



An example of the Sources and Analysis in NVivo

**Appendix V**

**Mock IELTS Test Observation Schedule**

**Name of Candidate Observed:** .....

**Date:** .....

<b>IELTS Writing Task One</b>													
1.	Time	Approximate time spent on understanding the question .....secs/mins											
		Time to start writing.....secs/mins											
		Total time spent on the task..... mins											
2.	Brainstorming												
3.	Planning												
4.	Revision												
5.	Pausing												
6.	Editing												
7.	Reference to the question												
8.	Making Drafts												
<b>IELTS Writing Task Two</b>													
1.	Time	Approximate time spent on understanding the question .....secs/mins											
		Time to start writing.....secs/mins											
		Total time spent on the task..... mins											
2.	Brainstorming												
3.	Planning												
4.	Revision												
5.	Pausing												
6.	Editing												
7.	Reference to the question												
8.	Making Drafts												