

Investigating the Predictive Validity of IELTS for a Teacher Education Program in UAE

Slim Khemakhem

A thesis submitted in partial fulfilment of the requirements
of the University of the West of England, Bristol
for the degree of Doctor of Philosophy

Presented to the Department of Humanities Languages and Social Sciences
University of the West of England
Bristol, UK
November 2016

Abstract

This research study investigates the predictive validity of IELTS test scores as a graduation requirement for a Teacher Education Program in the UAE. The argument is that the use of the IELTS cut-off score (band 6) as a predictor of post-graduation performance in schools may not be justifiable based on differences in context-related interaction and their impact on language proficiency. Subjects in the present study were given a mock IELTS speaking test, and were recorded while giving classes on teaching practicum to examine the correspondence between the two performances and the impact of each context on their lexical diversity scores.

A comparison between lexical diversity scores in each context using index D , revealed that there was no correlation between the two sets of scores as lexical diversity scores in the classroom were clearly lower than IELTS scores. Moreover, a comparison between the subjects' band-scores on the IELTS speaking test and their mentors' grades on teaching practicum revealed that there was no correlation. Corpus linguistic tools were also used in this study to compare content-word frequency lists produced in the IELTS speaking test with those produced in the classroom. Results showed that there were clear discrepancies, which indicated that context was a determining factor for the subjects' lexical choices. A further qualitative analysis took an emic perspective following a Conversation Analysis approach in analysing the subjects' talk in the classroom. Seedhouse's (2004) framework for the analysis of the architecture of classroom interaction and Walsh's (2011) combined Corpus Linguistics and Conversation Analysis (CLCA) approach were used as frames of reference. The analysis revealed that teacher repeats of lexical items in the classroom played a major role in reducing the subjects' lexical diversity scores due to their frequency and variety. A total of 16 types of teacher repeats was identified. Those types were classified into 6 main categories related to different features of classroom discourse such as control of interaction, asking questions, giving feedback, and so on. Consequently, the predictive validity of IELTS scores was found to be weak for the context of a teacher education program.

A two-step argument-based validation process is suggested at the end of this research study such that IELTS scores can be included as part of the requirement for the graduation of the Bachelor of Education program students. The first step is to adopt a complementary assessment based on

a rubric created for the purpose of assessing classroom-based English language proficiency (CBELP). It is suggested that this assessment is administered by a trained mentor during classroom-based observations conducted on teaching practicum. The second step is to formulate a composite score based on the minimum required band-scores for both the IELTS and the CBELP assessments. The interpretation of that score should be supported by a validating argument leading from score interpretation to related decisions regarding the graduation of Bachelor of Education students.

The thesis makes a contribution to theory in the fields of testing, lexical diversity and language proficiency, and furthers our understanding of classroom discourse.

Dedication

I dedicate this dissertation to the soul of my father Ali Khemakhem to whom I had promised to get a Ph.D. degree before he passed away, and to my mother Zaineb whose prayers kept me motivated all along the research journey. I also dedicate it to my wonderful wife Nadia -a patient, dependable, and supportive partner who was always there to take full responsibility of our family in my presence and in my absence. Last but not the least, I dedicate it to my three children Baya, Ali, and Aicha whose presence in my life kept inspiring me.

Acknowledgment

Before acknowledging people's contributions to this achievement, I wish to acknowledge my thankfulness to Allah the Almighty who gave me power, patience, and stamina to endure seven years of hard work and to manage to overcome hindrances of all kinds.

Many people in my life contributed in one way or another to this achievement. First, I would like to express sincere feelings of gratitude to my supervisor Dr Kate Beeching whose guidance, assistance, patience, and continuous support have been by all means exceptional. Second, I would like to thank Dr Guoxing Yu, Dr James Murphy, Dr Michael Daller, and Dr Jeanine Treffers-Daller for their assistance at different stages of the research and the writing of the thesis. I also extend special thanks to my large family including my siblings, my in-laws, and my friends in the UAE who kept encouraging me and who offered help to me and to my family in so many different ways.

Finally, I would like to thank my colleagues at Fujairah Women's College who volunteered their time and effort to conduct the mock IELTS speaking test for the subjects of the present research study. My heart-felt thankfulness also goes to my dear B.Ed. students who challenged their culture-related concerns about taking part in this study and who demonstrated full support throughout the process of data collection and subsequent stages.

May Allah bless you all.

Table of contents

Chapter 1: Introduction

1.1	Introduction.....	1
1.2	Theoretical background.....	2
	1.2.1 Language proficiency.....	2
	1.2.2 Language testing.....	4
	1.2.3 The context of the classroom.....	5
1.3	The study.....	7
	1.3.1 The main argument.....	7
	1.3.2 The main question.....	7
	1.3.3 The aim of the study.....	8
1.4	Thesis structure.....	9

Chapter 2: Language proficiency testing and classroom interaction

2.1	Introduction.....	12
2.2	Language proficiency.....	13
	2.2.1 Bachman & Plamer's model of language proficiency.....	13
	1- Language knowledge.....	14
	2- Strategic competence.....	15
	2.2.2 Cummins's model of language proficiency.....	17
	2.2.3 Canale & Swain's framework of communicative competence.....	19
	2.2.4 Audience design.....	21
	1- Accommodation theory.....	22
	2- Addressee design.....	24
	3- Intraspeaker and interspeaker variation.....	25
	2.2.5 English-for-Teaching.....	26
2.3	Lexical proficiency.....	28
	2.3.1 Lexical diversity.....	30

2.3.2	lexical diversity indices.....	32
2.4	Classroom interaction.....	35
2.4.1	The classroom context.....	35
2.4.2	The architecture of classroom interaction.....	38
2.4.3	Control of the interaction.....	40
2.4.4	Speech modification.....	41
2.4.5	Elicitation and questions.....	44
2.4.6	Repair.....	47
2.4.7	Scaffolding.....	49
2.4.8	Repetition.....	51
2.5	Language testing.....	52
2.5.1	An overview of the development of language testing.....	53
2.5.2	Bachman & Palmer's framework of language testing.....	54
	1- Characteristics of language use tasks.....	54
	2- Characteristics of language users and test takers.....	55
2.5.3	Testing speaking.....	57
2.5.4	The IELTS test	59
	1- The IELTS speaking test	60
	2- The development of the IELTS test	61
	3- The issue of language sampling in the IELTS test	63
2.6	Validity.....	64
2.6.1	Basic types of validity.....	65
	1- Criterion-related validity.....	66
	2- Content validity.....	66
	3- Construct validity.....	66
	4- Predictive validity	67
2.6.2	Recent types of validity.....	67
	1- Cognitive validity.....	67
	2- Context validity.....	78
2.6.3	Predictive validity.....	69

1-	Studies reporting weak predictive validity of the IELTS test.....	69
2-	Studies reporting strong predictive validity of IELTS test.....	72
3-	Using IELTS scores as a ‘hard criterion’ for success.....	73
2.6.4	Messick’s unitary framework of validity.....	75
2.6.5	Kane’s argument-based validity.....	78
2.7	Conclusion.....	80

Chapter 3: Methodology

3.1	Introduction.....	82
3.2	Focus of the study.....	82
3.3	Research questions.....	83
3.4	Research paradigm and epistemology	83
3.5	Methodological approach	85
3.6	Research design.....	86
3.7	Data collection instruments.....	87
3.7.1	The mock IELTS speaking test.....	88
3.7.2	The class recordings.....	89
3.8	Data collection procedures	89
3.8.1	The setting	89
3.8.2	Sampling and permissions	91
3.9	Data analysis procedures.....	92
3.9.1	Quantitative analysis tools.....	93
1-	CHAT and CLAN	93
2-	VOCD	96
3-	WordSmith tools	98
3.9.2	Qualitative analysis tools.....	99
1-	Seedhouse’s framework for the analysis of the interactional architecture of the L2 classroom.....	100
2-	Corpus linguistics.....	103
3-	Walsh’s combined approach (CLCA).....	104

3.10	Research validity and reliability	106
3.10.1	Validity	106
3.10.2	Reliability	109
3.11	Research ethics	110
3.12	Methodological issues	111
3.8	Conclusion.....	112

Chapter 4: Results

4.1	Introduction.....	114
4.2	The quantitative analysis.....	114
4.2.1	The correlation between IELTS and class D-scores.....	114
1-	Test of normality.....	117
2-	The mean.....	118
3-	The standard deviation.....	118
4-	Pearson correlation for D-scores of IELTS and class performances.....	119
4.2.2	The correlation between MCT grades and IELTS scores.....	119
4.3	Corpus linguistics analysis.....	121
4.3.1	A comparison between IELTS and class content-word frequency lists.....	122
4.3.2	An analysis of class content-word frequency list.....	123
1-	Verb lemmas.....	124
2-	Noun lemmas.....	127
3-	Adjective lemmas.....	128
4.4	The qualitative analysis.....	129
4.4.1	A taxonomy of teacher repeats.....	129
4.4.2	Interaction-control repeats.....	131
1-	Repeats in IRF exchanges.....	131
2-	Repeats in group work.....	132
3-	Repeats in repair phases.....	133
4.4.3	Question repeats.....	134

1-	Elicitation and modification repeats.....	134
2-	Strict question-repeats.....	135
3-	Think-time repeats.....	136
4.4.4	Feedback repeats.....	138
1-	Confirmation repeats.....	138
2-	Praise-word repeats.....	139
3-	Repair repeats.....	141
4.4.5	Key-word repeats.....	142
1-	Lesson key-word repeats.....	142
2-	Activity key-word repeats.....	145
3-	Story key-word repeats.....	147
4.4.6	Approach-related repeats.....	147
1-	Language-drill repeats.....	147
2-	Scaffolding repeats.....	149
4.4.7	Procedural repeats.....	151
1-	Instruction-clarification repeats.....	151
2-	Classroom management repeats.....	154
4.5	Conclusion.....	156

Chapter 5: Discussion of the findings

5.1	Introduction.....	159
5.2	Do student teachers display similar lexical diversity on the IELTS speaking test and the in the classroom?.....	160
5.2.1	The correspondence between language teaching and the IELTS speaking test.....	160
1-	Task-related cognitive demands and context-embeddedness.....	162
2-	The cognitive validity of the IELTS speaking task for a classroom teaching task ..	164
3-	The context validity of the IELTS speaking task for a classroom teaching task.....	165
5.2.2	The correspondence between the characteristics of the teacher trainee and the IELTS candidate.....	166
5.3	Do student teachers' scores on IELTS correlate with their teaching practicum grades?...	170
5.3.1	No correlation between IELTS scores and TP grades.....	170

5.3.2	The predictive validity of IELTS scores for an education program is weak.....	172
1-	The issue of language sampling.....	173
2-	The issue of inappropriateness of inferences.....	173
3-	The issue of generalisability of score interpretation.....	174
4-	The issue of test misinformed consequences.....	175
5-	The issue of using IELTS scores as a ‘hard criterion’	176
5.4	How does classroom interaction affect teachers’ lexical diversity scores?.....	177
5.4.1	The distinctive classroom context.....	177
5.4.2	The impact of lexical repeats.....	179
1-	Control-of-interaction repeats.....	179
2-	Question repeats.....	182
3-	Feedback repeats.....	183
4-	Key-word repeats.....	185
5-	Approach-related repeats.....	186
6-	Procedural repeats.....	187
5.5	Conclusion.....	190
 Chapter 6: Conclusion		
6.1	Introduction.....	192
6.2	Implications for practice.....	192
6.2.1	A complementary assessment for classroom-based language proficiency.....	193
1-	The design of the CBELP assessment.....	193
2-	Using the CBELP rubric.....	194
6.2.2	Establishing the predictive validity of the IELTS test and the CBELP assessment for a teacher education program	194
1-	Eliminating IELTS scores validity issues.....	196
2-	Developing an argument that validates IELTS and CBELP score-interpretation and ensuing decisions.....	197
6.3	Contribution to the current state of knowledge.....	201
6.3.1	The predictive validity of IELTS for a teacher education program.....	201
6.3.2	The variability of lexical diversity according to the context.....	202

6.3.3	A taxonomy of teacher repeats in the L2 classroom.....	203
6.3.4	A classroom-based English language proficiency assessment for teachers (CBELP).....	205
6.3.5	A corpus of classroom discourse and the mock IELTS speaking test for student teachers in the UAE.....	206
6.4	Limitations of the research study.....	207
6.5	Conclusion.....	208
References.....		210

Figures and tables

Figures

Figure 2.1	Bachman & Palmer’s framework of language ability.....	16
Figure 2.2	Cummins’s framework of language proficiency.....	18
Figure 2.3	The English-for-Teaching model.....	27
Figure 2.4	A three-way view of context.....	38
Figure 2.5	Correspondence between language use and language test performance	57
Figure 4.1	Proportions of content words in IELTS and in class.....	122
Figure 5.1	Cognitive demands and context-embeddedness in IELTS and class tasks.....	164
Figure 5.2	Convergence with the learners’ level of language proficiency.....	168
Figure 5.3	Convergence with the examiner’s level of language proficiency.....	169

Tables

Table 3.1	B.Ed. year-three Teaching Practicum assessment.....	91
Table 3.2	A concordance sample.....	98
Table 3.3	A word-list sample.....	98
Table 3.4	A key-word sample.....	99
Table 4.1	Student teachers’ D-scores for both IELTS and class	115
Table 4.2	Test of normality for the IELTS and class D-scores.....	118
Table 4.3	Pearson correlation of IELTS and class D-scores.....	119

Table 4.4	MCT grades and IELTS speaking test scores.....	120
Table 4.5	The test of normality for the MCT and IELTS scores.....	120
Table 4.6	Pearson correlation of MCT grades and IELTS speaking scores.....	121
Table 4.7	A summary of the rates of the top twenty content words in IELTS and in the class ..	122
Table 4.8	List of verbs in the top-twenty wordlist of the class.....	123
Table 4.9	List of nouns in the top-twenty wordlist of the class.....	127
Table 4.10	A taxonomy of repeat categories and types in the classroom.....	130
Table 6.1	An assessment rubric for classroom-based language proficiency.....	195
Table 6.2	An outline of the taxonomy of teacher repeats.....	204

Abbreviations

APA:	American Psychological Association
B.Ed.:	Bachelor of Education
BICS:	Basic Interpersonal Communicative Skills
CALP:	Cognitive Academic Language Proficiency
CA:	Conversation Analysis
CBELP:	Classroom-Based English Language Proficiency
CHILDES:	Child Language Data Exchange System
CHAT:	Codes for Human Analysis of Transcripts
CLAN:	Computerized Language Analysis
CL:	Corpus Linguistics
CLCA:	Corpus Linguistics & Conservation Analysis
EPTB:	English Proficiency Test Battery
ELT:	English Language Teaching
ELTS:	English Language Teaching in Schools
ESL:	English as a Second Language
ESP:	English for Specific Purposes
FWC:	Fujairah Women’s College
GPA:	Grade Point Average
HCT:	Higher Colleges of Technology
IELTS:	International English Language Testing System
IRF:	Initiation, Response & Feedback
LD:	Lexical Diversity
MCT:	Mentoring College Teacher
MST:	Mentoring School Teacher

MSTTR:	Mean Segmental Type-Token Ratio
SPSS:	Statistical Package for the Social Sciences
TOEFL:	Test Of English as a Foreign Language
TP:	Teaching Practicum
TTR:	Type Token Ratio
UAE:	United Arab Emirates
VOC-D:	Vocabulary Diversity
ZPD:	Zone of Proximal Development

List of Appendices

All appendices are available on an attached CD due to their size

Appendix 1 - HCT- B.Ed. Teaching Practice Competencies – Year 3

Appendix 2 - IELTS Mock Speaking Test 1

Appendix 3 - IELTS Mock Speaking Test 2

Appendix 4 - IELTS Mock Speaking Test 3

Appendix 5 - IELTS Mock Speaking Test 4

Appendix 6 - SPSS statistics for IELTS and CLASS lexical diversity (D-scores)

Appendix 7 - SPSS statistics for MCT grades and IELTS scores

Appendix 8 - Conversion table of MCT grades from a scale of 100 to a scale of 9 grades

Appendix 9 - A generic frequency list of the top-twenty content-words in IELTS and in the Class

Appendix 10 - T01 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 11 - T02 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 12 - T03 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 13 - T04 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 14 - T05 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 15 - T06 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 16 - T07 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 17 - T08 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 18 - T09 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 19 - T10 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 20 - T11 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 21 - T12- Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 22 - T13 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 23 -T14 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 24 - T15 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 25- T16 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 26 - T17 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 27 - T18 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 28 - T19 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 29 - T20 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 30 - T21 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 31 - T22 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 32 - T23 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 33 - T24 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 34 - T25 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 35 - T26 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 36 - T27 - Frequency list of the op-twenty content-words in IELTS and in the Class

Appendix 37 - Concordances of lemma *Write*

Appendix 38 - Concordances of lemma *go*

Appendix 39 - Concordances of lemma *come*

Appendix 40 - Concordances of lemma *sit*
Appendix 41 - Concordances of lemma *read*
Appendix 42 - Concordances of lemma *want*
Appendix 43 - Concordances of lemma *listen*
Appendix 44 - Concordances of lemma *look*
Appendix 45 - Concordances of lemma *see*
Appendix 46 - Concordances of lemma *finish*
Appendix 47 - Concordances of lemma *know*
Appendix 48 - Concordances of lemma *group*
Appendix 49 - Concordances of lemma *number*
Appendix 50 - Concordances of lemma *girl*
Appendix 51 - Concordances of lemma *sentence*
Appendix 52 - Concordances of lemma *excellent*
Appendix 53 - Concordances of lemma *good*
Appendix 54 - T01 Key-word list
Appendix 55 - T02 Key-word list
Appendix 56 - T03 Key-word list
Appendix 57 - T04 Key-word list
Appendix 58 - T05 Key-word list
Appendix 59 - T06 Key-word list
Appendix 60 - T07 Key-word list
Appendix 61 - T08 Key-word list
Appendix 62 - T09 Key-word list
Appendix 63 - T10 Key-word list
Appendix 64 - T11 Key-word list

Appendix 65 - T12 Key-word list
Appendix 66 - T13 Key-word list
Appendix 67 - T14 Key-word list
Appendix 68 - T15 Key-word list
Appendix 69 - T16 Key-word list
Appendix 70 - T17 Key-word list
Appendix 71 - T18 Key-word list
Appendix 72 - T19 Key-word list
Appendix 73 - T20 Key-word list
Appendix 74 - T21 Key-word list
Appendix 75 - T22 Key-word list
Appendix 76 - T23 Key-word list
Appendix 77 - T24 Key-word list
Appendix 78 - T25 Key-word list
Appendix 79 - T26 Key-word list
Appendix 80 - T27 Key-word list
Appendix 81 - T01 Class transcription
Appendix 82 - T02 Class transcription
Appendix 83 - T03 Class transcription
Appendix 84 - T04 Class transcription
Appendix 85 - T05 Class transcription
Appendix 86 - T06 Class transcription
Appendix 87 - T07 Class transcription
Appendix 88 - T08 Class transcription
Appendix 89 - T09 Class transcription

Appendix 90 - T10 Class transcription
Appendix 91 - T11 Class transcription
Appendix 92 - T12 Class transcription
Appendix 93 - T13 Class transcription
Appendix 94 - T14 Class transcription
Appendix 95 - T15 Class transcription
Appendix 96 - T16 Class transcription
Appendix 97 - T17 Class transcription
Appendix 98 - T18 Class transcription
Appendix 99 - T19 Class transcription
Appendix 100 - T20 Class transcription
Appendix 101 - T21 Class transcription
Appendix 102 - T22 Class transcription
Appendix 103 - T23 Class transcription
Appendix 104 - T24 Class transcription
Appendix 105 - T25 Class transcription
Appendix 106 - T26 Class transcription
Appendix 107 - T27 Class transcription
Appendix 108 - T01 IELTS speaking test transcription
Appendix 109 - T02 IELTS speaking test transcription
Appendix 110 - T03 IELTS speaking test transcription
Appendix 111 - T04 IELTS speaking test transcription
Appendix 112 - T05 IELTS speaking test transcription
Appendix 113 - T06 IELTS speaking test transcription
Appendix 114 - T07 IELTS speaking test transcription

Appendix 115 - T08 IELTS speaking test transcription
Appendix 116 - T09 IELTS speaking test transcription
Appendix 117 - T10 IELTS speaking test transcription
Appendix 118 - T11 IELTS speaking test transcription
Appendix 119 - T12 IELTS speaking test transcription
Appendix 120 - T13 IELTS speaking test transcription
Appendix 121 - T14 IELTS speaking test transcription
Appendix 122 - T15 IELTS speaking test transcription
Appendix 123 - T16 IELTS speaking test transcription
Appendix 124 - T17 IELTS speaking test transcription
Appendix 125 - T18 IELTS speaking test transcription
Appendix 126 - T19 IELTS speaking test transcription
Appendix 127 - T20 IELTS speaking test transcription
Appendix 128 - T21 IELTS speaking test transcription
Appendix 129 - T22 IELTS speaking test transcription
Appendix 130 - T23 IELTS speaking test transcription
Appendix 131 - T24 IELTS speaking test transcription
Appendix 132 - T25 IELTS speaking test transcription
Appendix 133 - T26 IELTS speaking test transcription
Appendix 134 - T27 IELTS speaking test transcription
Appendix 135 - Research Proposal and Ethic Clearance Form
Appendix 136 - Consent Form (English version)
Appendix 137 - Consent Form (Arabic version)

Chapter 1: Introduction

1.1 Introduction

The main argument of this research study has been motivated by an informal debate followed by a reflection on the graduation policy implemented by the Higher Colleges of Technology (henceforth HCT) in the United Arab Emirates for the graduation of the Bachelor of Education students (henceforth B.Ed. students). The policy sets a strict condition of achieving band 6 score in IELTS (International English Language Testing System) to be granted access to the final year of the program, and thereafter to graduate as teachers of English in the UAE schools. The initial reflection raised a number of questions about the rationale of such a condition and its appropriateness for high-stake decisions regarding the future of the B.Ed. students. Preliminary questions revolved around three main concepts and the relationships between them, namely language proficiency, language testing -with a focus on IELTS test, and classroom teaching. Some of those questions were: how do we define language proficiency in the classroom? Can it be gauged by a test, like IELTS? Can IELTS scores reflect teachers' language proficiency in the context of the classroom? In this research study, those relationships will be investigated with reference to the relevant literature and in light of empirical data to evaluate the predictive validity of the IELTS test scores for a teacher education program in the UAE. The study will contribute to the broader debate in applied linguistics over language proficiency and its relation with the context of language use. It will also contribute to the discipline of assessment and testing by suggesting an assessment tool that is specifically designed to gauge language proficiency in the context of the classroom.

This chapter will present a general and a brief theoretical background (1.2) for the three central concepts mentioned above with the aim of activating reflections related to the main focus of this study. It will also outline the study (1.3) by presenting the main argument, the aim of the research and the main question with its related sub-questions. Finally, an outline of the thesis structure (1.4) will be presented.

1.2 Theoretical background

1.2.1 Language proficiency

Notwithstanding the long history of language proficiency theories and their derivative approaches and methods, no unified definition has been agreed on due, first, to the complexity of the construct and, second to the heterogeneous viewpoints taken by theorists and specialists in language acquisition. Oller (1980, p.124) admits that defining language proficiency can be ‘one of the thorniest problems’ because it has a dual connection relating it to two different disciplines: the discipline of linguistics and the discipline of psychology. According to him, in order to try to define language proficiency, one has to be clear about the interfaces of both disciplines, and thereafter one has to decide on what should be included and what should be excluded. However, making such a decision is not an easy task, especially because the constructs of language and of intelligence intersect at a number of points, and it can be very difficult to disentangle one from the other in gauging an individual’s specific attributes. Bachman (1990, p.19) associates language proficiency with language ability but he confirms that it is difficult to give a precise definition and he describes any attempt to do so as ‘a complex undertaking’. Cummins (2000, p. 122) joins Oller and Bachman in asserting the difficulty of defining the construct and he describes the structure of language proficiency as ‘mythical’. He justifies his description by the existence of different variables that govern the relation between the different components of language proficiency such as context and language learners. He clarifies that there is ‘no one universal or absolute *structure of language proficiency* that can be identified across domains of use or experiences of learners’. Bialystok (2001, p.11) indicates that questions about the meaning of language proficiency despite their centrality to any research on language ability are ‘rarely if ever explicitly addressed’. She attributes the differences between the available definitions of language proficiency to differences in theoretical backgrounds or epistemologies. Theoreticians, according to Bialystok, refer to varied sets of assumptions about language and language learning, which create ‘disparate’ and sometimes conflicting perspectives.

Such agreement among theorists on the difficulty to formulate one definition of language proficiency creates some room for ambiguity and probably confusion in practice, especially when it comes to testing language proficiency. Practitioners in language testing can adopt one

perspective and ignore others which can make their practices and their inferences valid only from that perspective. The work done here is sympathetic to this position, and I seek to find common grounds between the different views of proficiency, particularly in the next chapter.

Initially, a one-dimensional view of language proficiency was dominant for a long period of time until componential views appeared and gained universal acknowledgment. Oller, who is one of the proponents of the unitary view of language proficiency, indicated that intelligence tests involved ‘a deep language factor of propositional reasoning which must be grammar-governed in some non-trivial sense’ (1980, p.128). However, this simplistic view of language proficiency does not account for the complexity of factors involved in interpreting a candidate’s score on a language proficiency test. There are different components that interact in a language proficiency test to give a picture of a candidate’s level. Even when successful research was conducted later on, and supported the view that a single factor was not sufficient to define language proficiency, Oller remained of the view that all language tests were characterized by one common underlying factor. In his comment on Bachman and Palmer’s (1980) research, Oller (1980, p.130) concedes that their two-trait model is carefully designed and uses ‘more sophisticated statistical methods’ that make their model superior to the one-trait model. However, he contends that the variance of factors in language proficiency tests demonstrated by the two-factor model is small and that represents a limitation of the significance of their findings. In other words, the added complexity brought to the model is not justified because of the limited explanatory power it provides, i.e. it is better to keep the model as straightforward as possible.

Despite Oller’s (1980) criticism, Bachman (1990, p.68) emphasizes that language proficiency consists of ‘several distinct but related constructs’ and provides a theoretical framework of communicative language ability that is based on three main components: language competence, strategic competence and psychophysiological mechanisms. Bachman and Palmer (1996) provide a very clear and well developed two-trait framework of language proficiency that has become a frame of reference for many linguists and researchers in language proficiency and language testing for the last two decades. In fact, along with some other models, it will be a major framework of reference in this research study as well for both language proficiency (2.2.1) and language testing (2.5.2).

Recent trends in research in language proficiency have focused their studies on lexical proficiency as a viable indicator of language proficiency. Their main argument is that lexis plays a major role in delivering meaning, and any failure in using words causes communication breakdown (Crossley et al., 2011). As a result, the last decade saw a clear growth in lexical studies and especially in lexical diversity indices. VOC-D has been proved to be a reliable index of lexical diversity that does not depend on the text length as it was the case of TTR which preceded it. In this study, lexical diversity and index VOC-D are going to be very instrumental in comparing language proficiency on the IELTS test and in the classroom. Further details will be presented in (2.3.1 & 2.3.2).

1.2.2 Language Testing

Davies (1990) indicates that Applied Linguistics developed three experimental procedures to test hypotheses, namely, language testing, second language acquisition studies, and discourse analysis. He further suggests that language testing is the most applied among the three as it is more concerned with ‘language demands’ and ‘selection requirements’. He also notes that over time language testing gained a theoretical power that developed at an even faster pace than its practical power, which made the field gain the status of a separate discipline. It developed qualitative measures alongside the quantitative measures for which it was known, and it widened its scope to cover evaluation of courses, materials, projects, and so on. That development moved ‘reactively’ between prioritizing reliability in the 1960s and 70s during the structuralist era which focused on form and accuracy to maximizing validity in the 1980s and 90s during the communicative era which focused on fluency and meaning. According to Davies (1990, p.77) the shift happened ‘from the typically reliable, structuralist, objective, discrete point tests of the 1960s into a more communicative mode’ in the 1980s.

It was Oller’s (1979) seminal work that marked that move into the communicative era of language testing through his discussion of integrative and pragmatic tests as opposed to discrete-item tests that were associated with the structuralist approach. Davies (1990) reports that Oller’s (1979) work on pragmatic language testing and expectancy grammar had an influence on the development of language proficiency tests at varying levels. Moreover, Bachman and Palmer’s (1996) framework focusing on language ability and on the correspondence between the

language use and the language test performance represents a landmark in the development of language testing theory in the communicative era.

Some of the international test batteries that underwent changes in their move towards a communicative mode included IELTS which adopted a ‘moderate’ approach that compromised between the structuralist and the communicative approaches with a focus on Bachman’s interactional ability. ‘IELTS represents a kind of regression to the mean, a (good) compromise between the extremes of the structural and the communicative’ (Davies 2008, pp.108-109). By looking at the history of the test, there is enough evidence provided by developers of the test and by published validation studies that it has been revised regularly to ensure that it is a valid test. However, the question that has not been addressed straightforwardly so far is whether IELTS scores are valid in any context. Though IELTS developers have never claimed that it is valid in every context, test users seem to have developed that assumption as it is the case in the present study. It is on this theme that this research study is focused, specifically seeking to answer the question: to what extent can scores obtained on such a ‘valid’ test be predictive of student teachers’ language proficiency in the specific context of the L2 classroom?

The following section on the context of the classroom will provide some background knowledge that will pave the way for a detailed review of the classroom interaction in chapter 2 (2.4) in order to discuss the predictive validity of IELTS scores for that specific context.

1.2.3 The context of the classroom

Sinclair and Coulthard (1975, p.15) indicate that interest in interaction in the classroom dates back to the 1940’s, but they note that models of interaction analysis were flawed with imperfections. They suggest their system of classroom interaction analysis in which they posit that classroom discourse consists of a number of exchanges. They identify *Initiation*, *Response*, and *Feedback* (IRF) as a basic three-part structure of an exchange in classroom interaction. This structure helped clarify the nature of communication in the classroom where the teacher initiates (I) a communicative exchange usually by asking questions and giving prompts; a student responds (R) to the teacher’s initiative; and the teacher gives feedback (F) on the student’s response. They identify eleven sub-categories of teaching exchanges in which variations of the IRF structure occur, such as [I (R)] when the teacher just informs (Sinclair & Coulthard 1975,

p. 52), and [I R F (I_b) R F] in re-initiation moves to help students who cannot get the right answer, or to choose another student (1975, p. 54), and so on.

Walsh (2011, 2013) indicates that IRF structure is the most commonly used structure of classroom discourse and that it helps to understand the nature of classroom interaction. It explains why teachers talk twice as much as their students.

In his ethnographic study of classroom behaviour which was based on a corpus obtained from videotaping and transcribing nine classes in an American school, Mehan (1979) identifies a structure of classroom interaction exchange that is similar to IRF exchange structure of Sinclair and Coulthard's (1975), but he calls it *Initiation, Reply, Evaluation* (IRE). However, the additional value that Mehan's model brings is the complete model of the hierarchical and sequential organization of lessons. Mehan (1979) notices that lessons are composed of three main phases: the opening phase to announce the lesson, the instructional phase to exchange academic information, and the closing phase to reformulate achieved learning and to move to other classroom activities. This is accomplished through the verbal and non-verbal behaviour of the participants which is organized into *interactional sequences*. *Directive* and *informative* sequences represent the opening and closing phases, and *elicitation* sequences represent the instructional phase.

Classroom interaction and teacher talk have also been researched in connection with the impact of approaches and methods of language teaching advocated by theorists and well-known practitioners. Proponents of the communicative approach have always called to regard the L2 classroom as an authentic context of language use like any other contexts with less teacher control of interaction through questions and error correction. Their call came as a reaction to what was described as unauthentic language use during the structuralist era that used language drills, recitals, repetitions, and so on through which the teacher controlled most of the interaction. Van Lier (1988, p.272) adopts a balanced approach in which he indicates that some forms of communication like choral drills, pattern repetitions, and dialogue recitals are helpful for learning because they are 'structured and orchestrated forms of interaction'. He adds that 'many classroom activities have strong ritual elements' which are conducted through 'specific rules, have repetitive or cyclical elements, and leave little room for variation or change' (Ibid).

Van Lier suggests ways of using those forms along with questions and repair in a more balanced approach that optimizes learners' share and reduces the teacher's control.

Walsh (2011, pp.24-25) defines classrooms as 'unique social contexts in their own right' due to a variety of interactional features that make them distinguished from other contexts. Teacher elicitation strategies, student responses and teacher's feedback, error correction, wait time and teacher questions are salient features that make the classroom a unique context of social interaction. Teacher talk and language proficiency are controlled by those features.

1.3 The study

1.3.1 The main argument

As mentioned in the introduction of this chapter, the setting of IELTS band-score 6 as a strict condition for student teachers at HCT to graduate as teachers of English needs to be justified. In particular, the question whether IELTS is an appropriate test to be taken by the B.Ed. students to gauge their language proficiency level for teaching English in the L2 classroom is an important one. Related to that is the idea that scores obtained on the test may not necessarily give accurate predictions of student-teachers' performance in the context of the classroom. The main argument is that IELTS test scores might not be good indicators of classroom-based language proficiency particularly based on the assumption that the social context of the classroom is very special that interaction on the test would not capture the main features of classroom interaction. Consequently, score interpretations and the high-stake decisions taken on the basis of relatively unjustified score-thresholds might not be fair ones. This thesis seeks to explore in detail the appropriateness of the IELTS scores for the specific gate-keeping purpose to which it is put in the UAE, i.e., who can be a teacher of English.

1.3.2 The main question

To reiterate, the main question of this research study is:

Can IELTS test scores make accurate predictions of student teachers' performance in the L2 classroom?

Three sub-questions are suggested to guide this research study in order answer the main question:

- 1) Do student teachers display similar lexical diversity on the IELTS speaking test and in the classroom?
- 2) Do IELTS scores match college mentors' grades for student teachers on teaching practicum?
- 3) How does classroom interaction affect student teachers' lexical diversity scores?

1.3.3 The aim of the study

The aim of this research study is to investigate the relationship between the IELTS test and classroom teaching to evaluate the predictive validity of IELTS test scores for a subsequent student-teacher performance in the classroom. In order to offer valid conclusions, the present research will focus only on the speaking component of the IELTS test as it is this aspect which characterises the teacher's language performance in the classroom, and which will inform decision making about graduation. Furthermore, the focus on the speaking skill will be narrowed down to a focus on the lexical diversity of the student teachers. The choice of lexical diversity as an indicator of language proficiency is justified by a number of research studies in applied linguistics that will be reviewed in chapter 2. The subjects' lexical diversity scores on the IELTS speaking test will be compared with their lexical diversity scores in the classroom and with the college mentors' grades on teaching practicum (henceforth TP) to draw statistical conclusions on the strength of the relationship between the test and the classroom performance. A qualitative analysis of classroom performance based on a corpus of the subjects' recorded classes will be conducted to identify classroom interaction features that impact student teachers' language proficiency in the classroom and their relevance to their performance on the speaking test of IELTS. The final conclusions will suggest other ways that validate score interpretations and decision-making based on IELTS test scores while taking into consideration the impact of classroom interaction on graduating teachers' language proficiency.

1.4 Thesis structure

In order to answer the research question, this thesis will be structured in the following way:

In chapter 2, the literature review gives a detailed background knowledge about the main concepts in this research study and their relevant components with summaries of pertinent research studies to serve as references in subsequent chapters. They are organized as follows:

- 1- The first section (2.2) focuses on models of language proficiency that are relevant to the main argument of this research study. They include Bachman and Palmer's model of language ability in (2.2.1), Cummins's model in (2.2.2), Canale and Swain's model of communicative competence in (2.2.3), Bell's model of audience design in (2.2.4), and Freeman et al. model of English-for-Teaching in (2.2.5).
- 2- The second section (2.3) reviews lexical proficiency, and it includes lexical diversity (2.3.1) and lexical diversity indices (2.3.2).
- 3- The third section (2.4) focuses on classroom interaction, and it includes the classroom context in (2.4.1), Seedhouse's architecture of classroom interaction in (2.4.2), control of the interaction in (2.4.3), speech modification in (2.4.4), elicitation techniques in (2.4.5), repair in (2.4.6), scaffolding in (2.4.7), and repetition in (2.4.8).
- 4- The fourth section (2.5) reviews briefly the development of language testing in (2.5.1) then Bachman and Palmer's conceptual framework of the correspondence between the language test task and the language use task in (2.5.2). Testing speaking is presented in (2.5.3) and the IELTS test in (2.5.4).
- 5- The Final section (2.6) reviews validity. Basic types of validity are reviewed in (2.6.1), and recent types are reviewed in (2.6.2). Then predictive validity is reviewed in (2.6.3) followed by Messick's unitary framework of validity in (2.6.4) and Kane's argument-based validity in (2.6.5).

In chapter 3, the methodology of the research study is presented. Section (3.2) presents the focus of the study, while section (3.3) presents the research questions. Section (3.4) discusses the research paradigm and epistemology. Section (3.5) elaborates on methodological approach, and section (3.6) presents the research design. Data collection instruments and data collection procedures are detailed in sections (3.7) and (3.8) respectively. Then, data analysis procedures

are discussed in section (3.9). Validity and reliability issues are discussed in section (3.10). A brief explanation of research ethics is given in section (3.11). Methodological issues that have arisen during the research are discussed in section (3.12), and a brief reflection on the research process is given in section (3.13).

In chapter 4, the results are presented. Section (4.2) provides the quantitative results which include the correlation between VOC-D values for IELTS and class performances in (4.2.1). Sub-section (4.2.2) presents the correlation between the Mentoring College Teacher's (henceforth MCT) grades and IELTS scores. Corpus linguistics analysis is presented in section (4.3) with a comparison between the statistics of word frequency lists for the class and the IELTS test in (4.3.1), and an analysis of class top-twenty content-word list in (4.3.2). Finally, section (4.4) presents the qualitative findings and reports 6 categories of teacher repeats: interaction-control repeats in (4.4.2), question repeats in (4.4.3), feedback repeats in (4.4.4), key-word repeats in (4.4.5), approach-related repeats in (4.4.6), and procedural repeats in (4.4.7).

In chapter 5, a discussion of the findings is presented. Section (5.1) presents a reminder of the main and the sub-questions of the study. Section (5.2) answers the first sub-question by discussing the correspondence between language teaching and the IELTS speaking test in (5.2.1), and the correspondence between the characteristics of the teacher trainee and the IELTS candidate in (5.2.2). Section (5.3) answers the second sub-question by commenting on the relationship between IELTS scores and the MCT grades in (5.3.1), and by evaluating the predictive validity of IELTS scores for a teacher education program (5.3.2). Section (5.4) answers the second sub-question by discussing the distinctiveness of the classroom context in (5.4.1), and the impact of lexical repeats in (5.4.2).

In chapter 6, concluding ideas are presented. Section (6.2) presents implications for theory and practice which includes a complementary assessment for classroom-based language proficiency (6.2.1), and establishing the predictive validity of IELTS and CBELP scores for a teacher education program (6.2.2). Section (6.3) presents contributions to the current state of knowledge which includes the predictive validity of IELTS scores as a pre-graduation requirement for a teacher education program in (6.3.1), the variability of lexical diversity according to the context of the language use in (6.3.2), a taxonomy of teacher repeats in the L2 classroom (6.3.3), a

classroom-based English language proficiency assessment (CBELP) in (6.3.4), and a corpus of classroom discourse and a mock IELTS speaking test (6.3.5). Finally section (6.4) presents the limitations of the present research study and implications for future research.

Chapter 2: Language proficiency testing and classroom interaction

2.1 Introduction

Language proficiency, lexical proficiency, classroom interaction, language testing, and validity are five focal areas in this chapter. The rationale for overviewing the literature in these areas is that they have a strong bearing on the main research question which investigates the adequacy of using IELTS scores to predict student teachers' classroom performance. First, language proficiency is discussed in section (2.2) as it represents the central construct around which the other subordinate concepts revolve. In this first section, five models of language proficiency that have strong connections with the main argument of this research are presented: (2.2.1) Bachman and Palmer's model of language ability, (2.2.2) Cummins's model of Cognitive Academic Language Proficiency, (2.2.3) Canale and Swain's model of communicative competence, (2.2.4) Bell's audience design, and (2.2.5) the model of Freeman et al. of English-for-Teaching.

The second section which is on lexical proficiency (2.3) provides a rationale based on the reviewed research literature for the choice of lexical proficiency to investigate language proficiency. In (2.3.1) the focus is narrowed down to lexical diversity as a measurable construct of lexical proficiency. Then, lexical diversity indices are discussed in (2.3.2), with a focus on index *D* as a reliable index of lexical diversity that will be used in this study .

The third section (2.4) focuses on classroom interaction. It reviews the classroom context (2.4.1), the architecture of classroom discourse (2.4.2) and six main features of classroom interaction, namely control of interaction (2.4.3), speech modification (2.4.4), elicitation (2.4.5), repair (2.4.6), scaffolding (2.4.7), and repetition (2.4.8).

The fourth section of this chapter focuses on language testing (2.5) which is first presented through a review of the main developments in the field in sub-section (2.5.1). Then, a special focus is given to Bachman and Palmer's model of language testing in sub-section (2.5.2) due to its importance in analysing the correspondence between the test task and the language use task. Sub-section (2.5.3) gives an overview of the development of IELTS.

The final section (2.6) reviews validity as it is the main concept that is investigated by this research study. Basic types of validity including criterion-related validity, content validity, construct validity, and predictive validity are briefly reviewed in sub-section (2.6.1). Sub-section (2.6.2) reviews recent types of validity including cognitive validity and context validity. Messick's (1989) unitary framework of validity is presented in sub-section (2.6.4) as a general frame of reference, and Kane's (1992) argument-based validity is discussed in sub-section (2.6.5) as a potential validation process for the IELTS scores in the context of the present study.

What is noteworthy in this chapter is that similar research studies are not presented in separate sections, but are integrated in the main above-listed sections according to their relevance to the main areas of focus.

2.2 Language proficiency

This section on language proficiency aims to provide a theoretical background that will help to understand the construct with reference to some theoretical models that have been consistently used as frames of reference, and that have a direct bearing on the main concepts discussed in this research. The choice of language proficiency models to be reviewed in this chapter depends mainly on their relevance to the main concepts suggested for investigation in this section on language proficiency, which are: language ability, the role of context and its impact on performance, the impact of communicative strategies on language performance, the role of accommodation strategies in language performance, and finally English language proficiency in the classroom.

2.2.1 Bachman & Palmer's model of language proficiency

Bachman and Palmer's model (1996) has been instrumental for many researchers in language proficiency and language testing because it presents a detailed and coherent framework that helps to understand the relationship between language proficiency and language tests. This framework will be used in this research as well to help discuss the central argument which is whether scores obtained on IELTS are good indicators of language proficiency for student

teachers in a classroom context. In a more precise way, whether lexical diversity scores displayed on the test are a true reflection of the subjects' lexical diversity in the classroom.

Language ability, which is identified as the most important characteristic of a language user is a pivotal concept in the context of the actual research study as it is measured by IELTS, and as it is used to make very important decisions as to whether a B.Ed. student is eligible to graduate as a teacher of English or not. Bachman and Palmer (2010, p. 43) define language ability as a 'construct' that represents the basis of an assessment task and related score interpretations. They consider that test takers are language users who interact with the characteristics of the test situation that can include other language users, and they use their language ability in different kinds of interactions while performing language use tasks. Therefore, Bachman and Palmer (2010, p.33) believe that test developers and test users need to understand language use to be able to generalize interpretations of a test taker's language ability to situations other than the test itself. In this research study, the subjects' language ability will be examined in light of Bachman and Palmer's (1996) framework and with reference to the subjects' performance on a test task (a mock IELTS speaking test), and on a language use task (a class teaching situation), to gauge the strength of correspondence between the two tasks and to be able to say whether the IELTS test scores are valid to make accurate inferences about the subjects' language performance in the L2 classroom.

Bachman & Palmer (1996, 2010) build their framework on Bachman's (1990) definition of language ability that involves two components: language knowledge and strategic competence (figure 2.1).

It is this combination of language knowledge and metacognitive strategies that provides language users with the ability, or capacity, to create and interpret discourse, either in responding to tasks on language tests or in non-test language use.

(Bachman & Palmer 1996, p.67)

1) Language knowledge

Language knowledge is defined as 'a domain of information in memory that is available for use by the metacognitive strategies in creating and interpreting discourse in language use' (Bachman

& Palmer 1996, p.67). It has two main components: organisational knowledge and pragmatic knowledge (Bachman & Palmer 2010, p. 45).

- 1- *Organisational knowledge* helps to interpret or produce utterances or sentences that are grammatically correct and that can be organized into oral or written texts. It is composed of grammatical knowledge (vocabulary, syntax, phonology/graphology) and textual knowledge (cohesion and rhetorical or conversational organisation).
- 2- *Pragmatic knowledge*, on the other hand, helps to create or understand discourse in relation with its meaning, the intentions of the language users and the setting. It is composed of functional knowledge (ideational functions, manipulative functions, heuristic functions, and imaginative functions) and sociolinguistic knowledge (genres, dialects/ varieties, registers, natural or idiomatic expressions, and cultural references and figures of speech).

This distinction between these two components of language knowledge is very important for the analysis of the correspondence between the language test task, which is the IELTS speaking test, and language use task, which is class teaching. For reasons that are going to be delineated in the section on lexical diversity (2.3.1), vocabulary knowledge is going to be taken as a benchmark of the subjects' organizational knowledge and is going to be used as a yardstick for the comparison between the test task and the language use task. Pragmatic knowledge is going to be examined through a qualitative analysis of the interaction between the subjects and their audience to account for differences in lexical diversity.

Bachman & Palmer's (1996, 2010) definition of pragmatic knowledge as a component of language ability is necessary to examine the subjects' intended meanings and relate them to their task contexts to show correspondence or lack of it, and the impact on the choice of vocabulary. Sensitivity to registers is going to be an important factor that determines the subjects' use of specific vocabulary in each situation.

2) *Strategic competence*

Strategic competence is defined as a set of 'higher order metacognitive strategies that provide a management function in language use as well as in other cognitive activities' (Bachman & Palmer 2010, p.48). *Goal-setting*, *appraising*, and *planning* are identified as three general areas within which metacognitive strategies operate (Bachman & Palmer 2010, p.49). Goal setting

refers to the goals that a language user intends to achieve when they engage in a communication act. Goal setting in the case of the IELTS test is presumed to be different from goal setting in a class situation. To impress the interviewer and to showcase language knowledge with all its components is of most importance in a speaking test situation for a test taker, whereas a successful implementation of a lesson plan and achieving planned goals is a priority for a teacher in the classroom situation. Appraising the communicative situation in a classroom and the required repairs to make sure that the students learn effectively is different from appraising in a test situation and the repair strategies that could be used to make sure that the interviewer is still ‘impressed’. Planning on a test depends very much more on the verbal or nonverbal feedback that an interviewee gets from the interviewer (though not much on IELTS), whereas planning in class depends on the students’ reactions and on the teacher’s appraisal of the teaching situation. Strategic competence will be revisited in section (2.2.3) which looks at the communicative competence framework of Canale and Swain (1981).

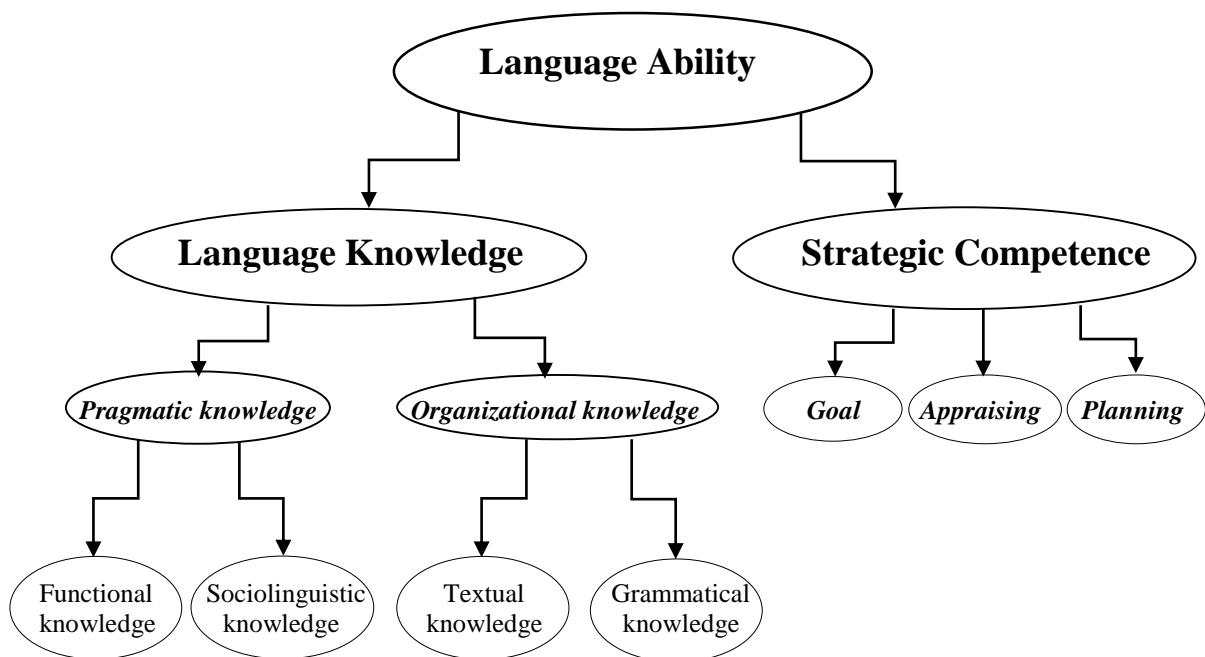


Figure 2.1 *Bachman & Palmer's framework of language ability*

2.2.2 Cummins's model of language proficiency

A second model of language proficiency that is worth considering in this research study is Cummins's model which will be referred to in analysing another dimension of the comparison between the test task and the language use task and its impact on the lexical choices made by the subjects in each context. Cummins (2000) postulates that context plays a major role in defining language proficiency. He contends that the complexity of cognitive demands on a language user depends primarily on the range of contextual cues available while carrying out a language task. In developing the main argument of the present study, I hypothesize that differences in context between the test and the classroom can cause the scores obtained on the test to be misleading when they are used to predict a teacher's performance in the classroom, because the classroom context is distinctive. In section (2.4.1), I refer to Sinclair and Coulthard (1975), Goodwin and Duranti (1992), Seedhouse (2004), and Walsh (2011) to explain why the classroom is a unique context and how it is unlikely that someone's score on the IELTS speaking test can reflect their performance in class. Cummins (2000) distinguishes between two types of proficiency, (1) conversational proficiency that he refers to as BICS (Basic Interpersonal Communicative Skills), and (2) academic proficiency that he refers to as CALP (Cognitive Academic Language Proficiency). He links CALP with registers of language associated with academic tasks and activities conducted in a school context, and he indicates that those tasks become more complex as students move up from one grade level to another. CALP is more important for us in this research because both the test task and the class teaching task are linked with academic contexts, though the range of lexical diversity differs according to the setting, the participants, and the activity. On the test and in order to achieve a high band-score, the subjects try to display their knowledge of language through diversification of vocabulary and use of different syntactic and discourse features, whereas in a teaching situation they usually try to simplify their language and rely on repetition in order to ensure that learning takes place. Even when they challenge their students with some new language items, they keep those items within the Zone of Proximal Development (ZPD) of the learners. To analyse and discuss the differences in support provided by the context and the effect on the subjects' lexical diversity, we need to refer to Cummins's (2000) framework (figure 2.2). He represents his two-dimensional framework in two intersecting continua, the vertical one represents the degree of

cognitive demands and the horizontal one represents the range of contextual support for any given language task.

Cummins's illustration of his framework (figure 2.2) situates language tasks and activities in four quadrants according to their context embeddedness and cognitive demands:

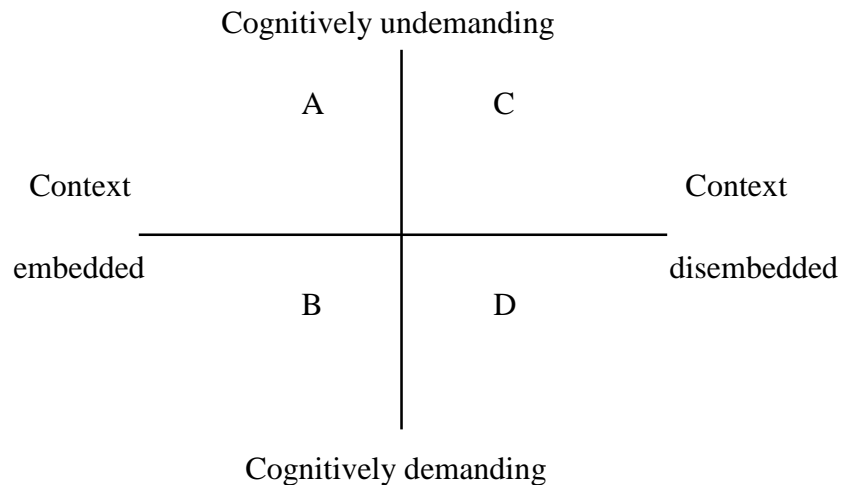


Figure 2.2 *Cummins's framework of language proficiency (2000, p.68)*

Used with permission of Multilingual Matters

- 1- *Quadrant A* is context embedded and cognitively undemanding. Cummins associates this with every day, outside the classroom, communication. The language is usually 'automatized' and supported by contextual cues. Formulating such language imposes very little cognitive pressure on the language user.
- 2- *Quadrant B* is context embedded but cognitively demanding as in persuasive speech or writing.
- 3- *Quadrant C* is context reduced but cognitively undemanding as in copying tasks or language drills.
- 4- *Quadrant D* is context reduced and cognitively demanding. Cummins associates this with mastery of academic functions as in writing essays.

Cummins's framework will be used to estimate the impact of context embeddedness or disembeddedness on the lexical choices made by the candidates on the test and in class situation.

2.2.3 Canale & Swain's framework of communicative competence

A third model that is going to be useful in the discussion of the findings in this research thesis is Canale and Swain's (1981) model which offers a different perspective focusing on the communicative competence of the subjects and the impact on their performance in both the test and the classroom context. Their framework offers to examine sociolinguistic factors emerging in each context, and the role they play in determining the lexical choices made by the subjects while performing their communicative acts.

Canale and Swain adopt Hymes's (1972) and Campbell and Wales's (1970) communicative competence which includes both grammatical competence and sociolinguistic competence. They distinguish between communicative competence which includes knowledge of grammar rules and language use, and communicative performance 'which is the realization of these competencies and their interaction in the actual production and comprehension of utterances (under general psychological constraints that are unique to performance)' (Canale and Swain 1981, p.6).

Canale and Swain (1981, p.16) indicate that Hymes defines communicative competence as the interaction between *grammatical competence*, *psycholinguistic competence*, *sociocultural competence*, and *probabilistic systems of competence*. The latter is depicted by Canale and Swain as being 'crucial' for second language learners to understand communication and to express themselves in a way similar to native speakers. In the case of the current study, probabilistic systems seem to play an important role in making the subjects' performance on the test different from their performance in class. As they strive to show that they are competent users of the language when they interact with a native or a near-native speaker while taking the test, their awareness of the rules of occurrences is at its highest level, whereas in a teaching situation their focus is on making sure their students learn effectively even by ignoring those rules and applying different rules such as the rule of repetition that can mismatch with probabilistic systems of competence.

In the absence of an integrative view of communicative competence, Canale and Swain (1981, p.20) define it as a 'synthesis of knowledge of basic grammatical principles, knowledge of how language is used in social contexts to perform communicative functions, and knowledge of how

utterances and communicative functions can be combined according to the principles of discourse.’ They comment that theories of communicative competence did not pay enough attention to communication strategies that speakers use in authentic situations which require knowledge of how to deal with breakdowns in communication or how to avoid grammatical forms that are not well mastered, or else how to talk with strangers when little is known about their social status. They purport that those strategies are important and they represent an integral part of communicative competence. They propose a theoretical framework for communicative competence that includes three main competences, namely *grammatical competence*, *sociolinguistic competence*, and *strategic competence* (1981, p.28). They consider that the relationship between an utterance and its social meaning varies across different sociocultural and discourse contexts. They postulate that participants engage in continuous evaluation and negotiation of social meaning. This is, as a matter of fact, one of the key concepts that the present research thesis is discussing. Can student teachers’ utterances on a speaking test be similar in terms of their lexical diversity to their utterances in a different sociocultural and discourse context which is the classroom? Do student teachers negotiate social meaning as they interact with a native or a near-native interviewer in the same way as when they interact with young second language learners in a classroom?

Canale and Swain assume that a theory of communicative competence interacts with human action and knowledge and is observable in communicative performance. They define *grammatical competence* as ‘knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology.’ (1981, p.29), and they indicate that it is important for any communicative approach that seeks to teach learners how to express meaning accurately.

Sociolinguistic competence includes sociocultural and discourse rules which are very important to interpret the social meaning of speech. Sociocultural rules define how appropriate utterances are with respect to the sociocultural context in which they are produced. This includes contextual factors like topic, role of participants, setting and norms of interaction. With respect to the present study, such contextual factors are important to consider in analysing the subjects’ performance on the test and in the classroom. Rules of discourse, on the other hand, include

cohesion and coherence with a focus on the combination of utterances and communicative functions rather than the grammaticality of utterances.

Strategic competence is composed of verbal and non-verbal communication strategies that are used to fix breakdowns in communication caused by ‘performance variables’ or lack of competence. Some of those strategies are associated with grammatical competence and some others with sociolinguistic competence. Student teachers in the case of the present study employ different communication strategies to achieve their communicative goals whether on the IELTS test or in the classroom. Studying some of those strategies will help to see if there is a correspondence between the test and the classroom contexts.

Like Hymes, Canale and Swain suggest that there is a rule of probability of occurrence that pervades each component, and that is essential for second language learners to know in order for them to achieve a ‘sufficient’ level of communicative competence. By rule of probability, Canale and Swain mean ‘the knowledge of relative frequencies of occurrence that a native speaker has’ (1981, p.31).

In language testing, Canale and Swain (1981, p.34) indicate that their framework of communicative competence suggests that language testing should provide language learners with opportunities to demonstrate their knowledge of a second language in a meaningful context (performance) rather than just by displaying knowledge about the language and its use in a paper-and-pencil test. Considering this view of language proficiency assessment, Canale and Swain (1981) do not seem to see a strong correspondence between performance on a test and performance in a classroom context. The analysis and the discussion of the findings of this research study will examine the strength of that correspondence in light of the communicative competence and with reference to the appropriateness of speech for the audience in each context. This will lead us to review the impact of the addressee on the speaker’s performance, which will be explained in the next section under Bell’s (1984) Audience Design.

2.2.4 Audience design

The main argument in this thesis suggests that there is a possible mismatch between what IELTS scores can indicate and their use to predict student teachers’ subsequent success as teachers of English in UAE schools. This argument rests on a hypothesis that student teachers’ performance

on the IELTS speaking test is different from their performance in a classroom situation due to a variety of factors, among which adjusting to the audience's level of proficiency is a principal one. In fact, addressing an examiner in a test situation suggests that the subjects optimize their lexical diversity to try to impress and to obtain higher scores. However, addressing L2 learners in a classroom suggests an opposite effect whereby they limit their lexical choices to avoid confusing the learners. This section presents Bell's (1984) audience design to provide a framework for the discussion of the differences in audience and the resulting impact on the lexical diversity of the subjects on a test and in a classroom situation.

Bell's 'audience design' suggests that the speakers' first concern in designing their speech is their hearers. Any differences within the speech of one speaker can be attributed to the influence of a second and a third hearer. According to Bell (1984), audience is not just the immediate addressee but it involves anyone who is present and participating in the act of communication actively or passively and from a close or a remote distance. In Bell's audience design, the first person is the speaker who is the primary participant responsible for designing and delivering speech to his/her audience. The audience involved in the communicative act is classified into three categories 'according to whether or not the persons are known, ratified, or addressed by the speaker' (Bell 1984, p. 159). In the first place comes the 'addressee' who is the central person in the audience as he/she is known to the speaker, ratified and addressed directly. In the second place comes the 'auditor' who is the second important person in the audience and who is known, ratified but not directly addressed by the speaker, and in the third place come two categories of audience, namely the 'overhearer' who is known to be there but is not ratified and is not addressed, and the 'eavesdropper' who is not known, not ratified and not addressed.

1) Accommodation theory

To justify the salience of the addressee in designing the speaker's speech, Bell (1984) refers to research studies and experiments conducted by Giles and his associates (1975, 1979 & 1980) and their notion of accommodation. Accommodation theory rests on the premise that speakers try to accommodate their speech to their addressees so that they reach their communicative goal out of the speech act they perform. Usually, speakers accommodate their speech through an act of convergence whereby they adapt their speech style so that it comes closer to that of their addressees' in order to get their approval. This accommodation factor is important for the

discussion of the main argument of this study which suggests that the level of lexical diversity on IELTS test is different than the level in a classroom situation due to the subjects' use of different accommodation strategies that theoretically require a higher lexical profile to adapt to an examiner's style on the test, but a lower profile to adapt to the learners' style in a classroom.

In a study that confirms the impact of accommodation on the lexical richness of the teachers in the classroom, Meara, Lightbown and Halter (1997, p.29) explore the lexical richness of the speech of ten teachers in ESL classes in Quebec, Canada, and its impact on vocabulary learning. In their analysis of samples of 500-word tokens, the researchers focused on lemmas of word types while using four levels of word frequency based on Nation's word frequency list (1986). The results revealed that all 10 teachers were performing at similar levels of lexical richness. About one third of the lemmas (36%) represented the basic level of richness which comprises about 500 high-frequency words like articles, prepositions, common greetings, etc., and around half of the total number of lemmas (53%) represented the second level which corresponds to Nation's 1000 most frequent words. Only 7% accounted for Nation's second 1000 most frequent words, and only one or two lemmas accounted for the fourth level that indicates that the students whose first language was French were ready to start their post-secondary education in English. The remaining 3% represented the rate of unusual words which were, for the researchers, an indication of lexical richness. That low rate made the researchers draw an initial conclusion that lexical richness for those classes was at a low level, but by considering that those classes were part of a five-month intensive ESL program (5 classes a day/ 5 days a week), they conclude that it could be rich enough. However, these findings and the initial conclusion that Meara, Lightbown and Halter (1997) reached is significant for the context of the present study because ESL classes are given on the average of one class a day for five days per week. Therefore, their study confirms that accommodation impacts teachers' lexical richness in the classroom which may not reflect their lexical richness outside the classroom, and specifically when they take a formal speaking test like IELTS.

In a follow-up study conducted by Horst (2009) within the same context but with larger corpus of teacher talk that was collected from three teachers on four distanced occasions, she confirmed the findings of Meara, Lightbown and Halter (1997). She found that teachers used 600 different word families with around 100 unusual words (off-list words). She also noted that lexical

richness was higher in activities when teachers were talking about language (135 off-list types), whereas in activities that included classroom management and read-aloud, off-list word types went down to 84 and 87 types. However, talk about the language represents only 17% of the total teacher talk, while 74% is taken up by ‘impoverished’ talk focusing on classroom procedures (Horst 2009, p.60).

2) Addressee design

By referring to the findings of a number of studies (Bikerton 1980, Douglas-Cowie 1978, Thelander 1982, Russel 1982, Trudgill 1981, and Coupland 1984) which compared the performance of informants on formal interviews or language used with a stranger to that used with peers, Bell (1984) indicates that the remarkable consistency of the results submitted by the studies that covered four different languages leave us with no doubt that there are clear differences between the speakers’ styles in an interview and their styles with peers. He then concludes that the effect of the addressee on the speaker’s style is a very strong one. He also signals that the findings of those studies revealed that accommodation of style does not take place with just one addressee but can happen with multiple addressees. As the speaker tries to converge to meet the addressees’ different styles, the degree of style shift will depend on the class of the addressee in relation to the speaker. In the case of the present study, the degree of style shift and its impact on the lexical diversity of the subject is expected to be a significant one due to clear differences in the class of the addressees.

Another factor that plays an important role in addressee design is the relation between the speaker and the addressee. In a situation where the speaker and the addressee do not know each other as in most interviews, the relation is low and the status of each one is high, which would affect the linguistic choices that would look more formal and standard as is the case of the relation between the subjects in this study and the IELTS examiner. On the other hand, when the relation between the two participants is strong as in friendship, the relation is high and the status is low, which would make the linguistic choices more informal and less standard as is the case between the subjects and their students in the classroom.

In the case of the present research study, Bell’s audience design is potentially a convenient framework that will help to understand the potential discrepancies in the lexical diversity of the

student teachers in two different contexts. The change in audience type engenders the use of accommodation strategies to be able to obtain the approval of the two different types of audience. The concept of convergence will help to demonstrate the need for the student teachers to raise their lexical diversity to approximate that of the examiner, to try to converge, whereas in the classroom situation student teachers need to match their lexical diversity to the learners' level of proficiency which requires the use and the repetition of high frequency words.

3) Intraspeaker and interspeaker variation

In his audience design framework, Bell (1984) distinguishes between 'intraspeaker' and 'interspeaker' variation in language style, and he posits that the relation between them is a relation of derivation. Intraspeaker variation is the style shift in a speaker's speech that is caused by extralinguistic factors like the addressee, the topic and the setting. Interspeaker variation relates to linguistic differences between speakers caused by differences in class, gender, social status, and so on. Bell (1984) expands on Mahl's (1972) finding that gave evidence that a speaker's awareness of their addressee is more important than their ability to pay attention to their speech, to indicate that the speaker's ability to pay attention to their speech and monitor their levels of formality (intraspeaker) is a result of their ability to see their addressees and decide on the level of required formality (interspeaker). For this reason, Bell postulates that intraspeaker variation is derived from and reflects interspeaker variation. He calls this the Style Axiom which states that 'variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension' (1984, p.151). This variation of style can give us different language performances in different contexts that reflect language proficiency in totally different ways. In the case of the present study, it is assumed that variation of style can give us two different language performances of the student teachers, one is appropriate for the context of the IELTS test that relies on diversifying vocabulary, and the other one is appropriate for the L2 classroom that relies on restricting vocabulary choices. As a result, we get two different profiles of English language proficiency, one that represents English-for-academic purposes displayed on the IELTS speaking test, and one for teaching and learning purposes displayed in the classroom. The next sub-section discusses English-for-teaching as discussed by Freeman et al. (2015).

2.2.5 English-for-Teaching

Freeman et al. (2015) define English-for-Teaching as a construct that ‘identifies relevant language skills the teacher can draw on to carry out instructional routines’ (Freeman et al. 2015, p.133). According to them, those classroom instructional routines, such as writing homework on the board, introducing activities, asking questions, giving feedback, and so on, ‘account for the English language that both appears in the curriculum content and is used as the medium through which that content is taught’ (2015, p.134). By analysing ten national ELT curricula and drawing on classroom research and on classroom language data from different countries, they identify three main areas in which English-for-Teaching operates. They are:

- 1- Classroom management
- 2- Understanding and communicating lesson content
- 3- Assessing students and giving feedback

Freeman et al. (2015) make a clear distinction between general language proficiency and context-specific classroom language which they call ‘English-for-Teaching’, a concept which has its background in ESP (English for Specific Purposes). They argue that the general assumption is that English teachers should have good command of English language to ensure good teaching and therefore good learning, whereas the issue is deeper than that. It is more about ‘connecting teachers’ general language proficiency with their familiarity and knowledge of classroom practices’ (2015, p.130). While some teachers might have good teaching skills they may not have good knowledge of English language to make use of those skills in an appropriate way. This argument seems very congruent with the argument of the present study. The main argument in this study puts the predictive validity of IELTS scores into question based on the assumption that what student teachers need to demonstrate on the test is language proficiency that is appropriate for the classroom and not general or academic language proficiency. Freeman et al. (2015) develop a new construct that they call ‘English-for-teaching’ which addresses the specific demands of language use in the classroom. It is both a language and a knowledge construct that reflects the dual role of English in the classroom which functions both as the means and the object of instruction. They claim that their model is derived from ESP, and they define English-for-teaching as:

The essential English language skills a teacher needs to be able to prepare and enact the lesson in a standardized (usually national) curriculum in English in a way that is recognizable and understandable to other speakers of the language.

(Freeman et al., 2015, p.132)

They identify three main elements in this definition:

- 1- Distinguishing between preparing a lesson and enacting it (pre-active versus interactive decisions).
- 2- The language and its underlying pedagogical focus are bounded by the curriculum
- 3- Comprehensibility of this English-for-teaching for other English language users in similar contexts.

Freeman et al. (2015, p.134) postulate that the construct reconciles and integrates ‘essential language’ which can be common to all instructional settings (global) and the particular use of language for classroom instruction (local).

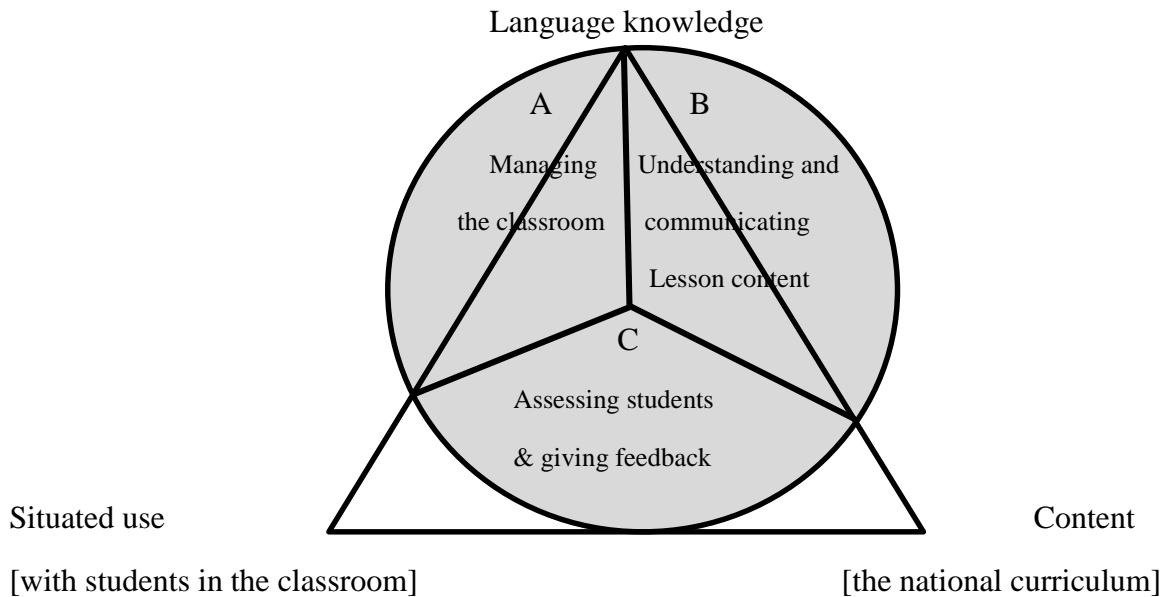


Figure 2.3 *The English-for-Teaching model (Freeman et al. 2015, p.135)*

Used with permission of Oxford University Press

2.3 Lexical proficiency

Lexical proficiency in this research study is taken as a viable indicator of language proficiency and is used as a yardstick to compare the oral performance of the student teachers on the language test (IELTS) with their performance in class in order for the researcher to draw conclusions on the ability of the test scores to make accurate predictions of performance in class. This choice is justified by the findings of recent research that has demonstrated a strong relationship between lexical competence and language proficiency (Laufer & Goldstein, 2004; Alderson, 2005; Albrechtsen, Hasstrup & Henriksen, 2008). Zareva, Schwanenflugel & Nikolova (2005) conducted a research study on a group of subjects including native speakers, advanced second language learners, and intermediate learners of English to find out if the relation between their lexical proficiency and language proficiency is a strong one. Their research revealed that vocabulary size, word frequency, word associations, and participants' associative domain within the same group are clearly sensitive to increasing scores on a language proficiency test that they designed. They concluded that the quality and the quantity of second language lexical competence increase as proficiency increases. Therefore, lexical proficiency can be taken as a credible indicator of language proficiency.

In the introduction to her study of language dominance among bilinguals and L2 learners, Treffers-Daller (2009, p.76) lists a number of reasons why it is important to focus on lexicon:

- 1- It plays a fundamental role in the latest versions of generative grammar as it determines the grammar, morphology and phonology that a speaker uses.
- 2- Psycholinguistic research focuses on lexicon in production and reception rather than on syntactic structures
- 3- There is important variability in the number of words that language users know and their knowledge about those words, because lexical knowledge depends on sociolinguistic variables.

Crossley, Salsbury, McNamara & Jarvis (2011, p.562), in their research study on the viability of computational indices in predicting human evaluations of second language learners' lexical proficiency, give three main reasons why lexical proficiency is 'of crucial interest to language acquisition and linguistic competence'. First, it can cause communication problems when

lexical items produced by learners are misunderstood. In fact, in the case of the actual study and in line with the previous argument, the subjects try to rely more on their choice of appropriate vocabulary to compensate for potential gaps and ambiguities in their speech due to their limited grammatical resources, especially in a test situation. In the same way, in a class teaching situation the subjects would be aware of the difficulties that their students face with English structures and patterns, so they put more focus on vocabulary to keep their instruction clear. Second, it is a good indicator of academic achievement. In fact, the subjects in the actual study know that a test situation is a perfect opportunity for them to display their knowledge of a wide range of academic vocabulary as a good indicator of their achievement. They also know that their limitations in grammatical accuracy, which is another criterion of the IELTS assessment rubric, could be compensated by showcasing lexical diversity. Third, it helps researchers to gain a better understanding of the way learners process and produce language. This third argument underpins the methodology of the current study as it uses lexical proficiency tools to investigate the variance of the subjects' lexical proficiency in different contexts and the validity of using the scores of one performance in a specific context to predict another one in a different context. A test situation would elicit lexical proficiency as it would entice the subjects to display their knowledge of a wide range of vocabulary, while a class teaching situation would elicit communicative skills and would incite the subjects to limit their lexical range to their students' ZPD in order to be able to conduct repair and scaffolding work. Corssley et al. (2011) survey a number of studies that cover different features of lexical proficiency (breadth of lexical knowledge, depth of lexical knowledge, and access to core lexical items) and they conclude that those features are good indicators of lexical growth and of second language learners' proficiency.

In a similar study that focuses on the same components of lexical proficiency, Crossley, Salsbury & McNamara (2012) investigate the viability of lexical indices in classifying texts written by second language learners of different language proficiency levels, and they find that word frequency and lexical diversity (using indices of word breadth) as well as word imaginability and word familiarity (using indices of access to core lexical items) are the strongest lexical predictors of language proficiency. In this research study, the focus is on indices of lexical diversity and word frequency to compare the language proficiency of the same subjects in two different contexts.

2.3.1 Lexical diversity

Recent research studies in language acquisition have displayed an increasing interest in lexical diversity as an indicator of lexical proficiency and subsequently as an indicator of language proficiency (Daller, Van Hoot & Treffers-Daller, 2003; Read and Nation 2006; Daller & Xue 2007; Yu, 2009; Treffers-Daller, 2009). Read's (2000) multi-dimensional model of lexical richness has been widely endorsed by researchers for lexical diversity measurement. He suggests four dimensions which are lexical variation (the percentage of different words), lexical sophistication (the percentage of low-frequency words), lexical density (the percentage of content words), and number of errors. Jarvis (2013, p.88) indicates that in recent research the terms 'lexical diversity', 'lexical variability', 'lexical variation', and 'lexical variety' are taken as synonyms as they are all used to mean the size of words in a language sample that are not repeated. As a result, lexical diversity is taken as the inverse of the rate of word repetition. He, also, indicates that the term 'lexical richness' was originally used to refer to the size of the mental lexicon but recently it has been used interchangeably with 'lexical diversity'. Although Jarvis highlights the problem of the indefinite use of terms and unclear definition of the construct, he points out that,

Research involving lexical measures has produced valuable findings concerning how learners' word choices contribute to the complexity and quality of their language use, and it has also shown that such measures serve as useful indices of learners' levels of language proficiency and stages of acquisition. (Jarvis 2013, p.89)

In this research study, lexical diversity is used to refer to the vocabulary size of the subjects which is measured by means of index D (it will be introduced in the next sub-section) in two different contexts to see if there is a correlation between the two performances that would help to tell whether performance on IELTS can predict performance in a class teaching situation.

Yu's (2009) research study on the differences in the performance of the same students on two different test components (speaking and writing), in two different situations, different times, and on two different tasks (format and topics), confirms the strength of lexical diversity as an indicator of language proficiency and the usefulness of index D as a valid measurement tool. He finds that despite the different situations and the different tasks, the lexical diversity D of

the candidates' performances on both the interview and the writing task is approximately at the same level. In fact, Yu's finding validates the use of lexical diversity and index *D* in the present research study as it is applied for two different situations (taking IELTS speaking test and teaching in the classroom), and two different tasks (answering questions of an interviewer and teaching young learners). This, in my opinion, adds value to the present study especially that *D* is not only computed for the performance of the same subjects on two different tasks but also for two different types of audience. Yu's (2009) research study is a post-test validation study of 200 compositions and 25 interviews of candidates who took the MELAB (Michigan English Language Assessment Battery) test between 2004 and 2005. He finds that the index of lexical diversity *D* is a good predictor of the overall quality of the compositions as the obtained statistics revealed a strong relation between the scores assigned to them and the lexical diversity scores. In his analysis of the relation between the lexical diversity scores and the scores assigned to the candidates for their performance on the interviews, Yu (2009) finds a significant relation and he indicates that *D* explains the larger variances of the overall quality scores of the interviews whereas other indices do not. Moreover, Yu finds that the 25 scripts of the interviews correlate significantly with the compositions of the same candidates, and he finds that the lexical diversity is at the same level despite the differences in contexts and in the type of the tasks. He, then, concludes that it is a good evidence of the validity of lexical diversity as a quality indicator, and of the usefulness of index *D* as a lexical diversity measurement tool for both speaking and writing performances.

In the present study, speech accommodation is regarded as an important factor that accounts for differences in lexical diversity between the subjects' performance on the mock IELTS speaking test and their performance in the classroom. Malvern & Richards' (2002) study offers a strong evidence on the strong link between accommodation and lexical diversity by using the same index of lexical diversity that is used in the actual study which is *D*. Malvern & Richards' study is a follow up on a research study that they conducted in 2000 in which they investigated different variables of speech accommodation in non-native teacher-interviewers and students. In that study, they found that lexical diversity was the most significant variable that marked teacher-interviewer accommodation in 34 audio-recorded interviews of British secondary school students who took an oral exam in French for the General Certificate of Secondary Education (GCSE). Malvern & Richards examine in depth lexical diversity as a major feature

of teacher-interviewer accommodation by using a ‘new’ measure of lexical diversity D after identifying problems with MSTTR (Mean Segmental Type-Token Ratio) index that was used in the initial study. They find that the new measure D overcomes problems identified with previous indices of lexical diversity, especially in accounting for the different sample sizes. They also find that teacher-interviewers accommodate their language to the language ability of the of the class as a whole and not of the individual students. Interestingly, the comparison between the average D of teachers and that of students reveal that teachers have lower lexical diversity and less variance than their students, which suggests that there is over-accommodation. Malvern and Richards note that this finding should not be attributed to the fact that teachers are giving the floor to the students to talk . They also note that there is no evidence that teachers use greater lexical diversity with students who have higher levels of language proficiency. This finding can be a supportive argument of the hypothesis of the present study that the subjects accommodate their lexical diversity to the level of their students in class which doesn’t reflect their real lexical diversity values that they display on the speaking test with an ‘expert’ interviewer.

2.3.2 Lexical diversity indices

Over the last fifteen years or so, computer-assisted research has brought rapid and significant developments in measuring lexical diversity which has helped to overcome the flaws of the traditional and widely-used Type-Token Ratio (TTR). Despite the fact that it has been under constant criticism for being vulnerable to text length effect, TTR remained in use for almost a century. It follows a simple formula which divides the number of types (number of different words) by the number of tokens (the total number of words in a text or a corpus). However, this formula does not work in a consistent way across all text lengths. Shorter texts usually tend to have higher rates of lexical diversity as text producers keep varying their vocabulary items, whereas longer texts tend to have lower rates or falling curves because text producers reach a point where they exhaust their lexical resources and start to reuse some of their vocabulary items as the text progresses. This recycling of used vocabulary increases tokens at the expense of types and culminates in a falling curve (Malvern & Richards, 1997; Malvern, Richards, Chipere & Dúran, 2004). Vermeer (2000) reviews different indices of lexical diversity and indicates that attempts to correct TTR by making some mathematical transformations in order to account for

text length, such as the index of Guiraud (1960) and ‘Corrected TTR’ of Carroll (1964) managed to reduce the problem but did not solve it. He concludes that TTR ‘is the worst measure of lexical richness’ (2000, p.69) and he expresses his astonishment that it was still widely used. Jarvis (2002) highlights three ‘widely used’ transformations of TTR, (1) Herdan’s index, (2) Guiraud’s index, and (3) Uber index, but he indicates that they either over-adjust or under-adjust TTR for the text length which makes their results erratic. A yet more recent attempt by Daller et al. (2003) to improve Guiraud’s mathematical formula, that they called ‘Advanced Guiraud’, proved to be unsuccessful as the same problem of sensitivity persisted (Malvern & Richards, 2009). It wasn’t until Malvern & Richards (1997, 2002), and Malvern, Richards, Chipere & Dúran (2004) published their new index of lexical diversity D and gave evidence that it was functioning accurately across texts of different lengths that researchers started to ignore TTR or use it just to demonstrate that it is a flawed index (Tidball & Treffers-Daller, 2007).

In this research study, TTR was initially used along with index D in the pilot stage but the inconsistent results yielded by TTR gave clear evidence that it was not useful for the comparison of two different performances that had different lengths (IELTS time range is clearly smaller than class time range), hence the decision was made to discard TTR and use index D for the computation of the subjects’ lexical diversity values.

The effect of using D should be to allow the comparison of two speakers on the basis of the number of types and tokens they produce irrespective of the length of a text or utterance (Daller & Xue 2007, p.152).

As a result of the clear flaws of TTR and its derivatives, Malvern & Richards (1997, 2002), McKee, Malvern & Richards (2000), and Malvern, Richards, Chipere & Dúran (2004) suggest index D as a measure of lexical richness that does not make transformations to TTR formula but creates a new model based on the calculation of the falling curve of TTR. It analyses the probability of new vocabulary being introduced into longer texts and generates a model of the way TTR varies with the number of tokens. Then, the mathematical model is compared with the empirical data to provide a measure of lexical diversity D . (McKee, Malvern & Richards, 2000)

VOCD is the program that was created specifically to compute D for transcripts processed by CLAN (Computerized Language Analysis) on files which are coded according to CHAT (Codes

for Human Analysis of Transcripts) as part of MacWhinney's (2000) CHILDES (Child Language Data Exchange System). *VOCD* computes D in the following way:

Vocd uses random sampling without replacement of tokens as the default mode for plotting the curve of TTR against increasing token size for the transcript under investigation.(...) each point on the curve is calculated from averaging the TTRs of 100 trials on sub-samples consisting of the number of tokens for that point, drawn at random from throughout the transcripts. (Malvern et al. 2004, p.56).

The default number of tokens that is used to draw the theoretical curve ranges between 35 and 50. Then, *VOCD* tries to find the best fit between the theoretical curves and the empirical one by adjusting the value of D until a perfect match is found between the actual curve and the closest curve drawn by the mathematical model. A high D value is an indicator of high lexical diversity and a low D is an indicator of low diversity.

Malvern et al. (2004) claim that the validity and reliability of D values is confirmed through extensive testing. Their claim was partly confirmed by Malvern & Richards (2002), Jarvis (2002), Tidball & Treffers-Daller (2007), Daller & Xue (2007), Treffers-Daller (2009), Yu (2009). It was found to perform better than any other index, but it was clearly criticized by McCarthy & Jarvis (2007) and Van Hout & Vermeer (2007) who recommended further refinements to achieve better accuracy.

McKee, Malvern & Richards (2000) specify three advantages of index D :

- 1- It does not depend on the number of words in the text.
- 2- It uses all the available data.
- 3- It does not represent one value of TTR but shows how it changes over a range of text lengths.

Considering these advantages and the fact that it is the best available tool that approximates real lexical diversity values, index D is used in this study to give a quantitative representation of the subjects' lexical diversity on the IELTS speaking test and in a classroom situation. The D -scores are used to compare the subjects' performances in the two contexts in order to see how much they correlate, and therefore to evaluate the extent to which IELTS scores can predict the

subjects' performances in the classroom. What is at stake here is the extent to which a *D*-score on IELTS can successfully predict a *D*-score in the classroom. If these two are unrelated, then it would appear that IELTS score is not a valid measure of an individual's ability to perform successfully in the classroom. Hence, the decision to use IELTS scores to qualify or disqualify candidates for a teaching job becomes questionable.

2.4 Classroom interaction

With reference to the British researcher Douglas Barnes, Cazden (1988) defines classroom discourse as the study of the communication system that a teacher sets up in order for the students to relate, through speech processing, what they already know with what their teacher presents to them. Classroom communication is a highly complex concept in a language class because language is at the same time the means and the goal of instruction. It is also central to all classroom activities due to its immediate effect on learning (Walsh, 2011). In this section, I am focusing on classroom interaction to gain a better understanding of student teachers' monitoring of classroom discourse and the subsequent impact on their oral performance, which -for the purpose of this research study- will be focused on their lexical diversity. That understanding will serve to see where the teacher's talk is governed by features of interaction that are specific of the L2 classroom, and that have a limiting role on the lexical choice made by the subjects in this study. In the next sub-section (2.4.1), I will discuss the classroom context and the factors that make it distinctive for the discussion of discourse. Then, in section (2.4.2) I will present the architecture of interaction in that context as suggested by Seedhouse (2004). After that I will review six main classroom interaction features that pervade that architecture. They include control of interaction (2.4.3), speech modification (2.4.4), elicitation (2.4.5), repair (2.4.6), scaffolding (2.4.7), and repetition (2.4.8). These features will be reviewed in view of their potential effect on the lexical diversity of the subjects and the impact on the meaningfulness of IELTS scores for performance in an L2 classroom context.

2.4.1 The classroom context

Sinclair & Coulthard (1975, p.12) emphasize the fundamental role of context in extracting meaning 'even when the most elaborated grammatical structures and most specific lexical items

are employed'. Citing Katz and Fodor's work on semantics (1963), they indicate that 'insurmountable problems' can be faced if we try to define lexical meaning independently from context. Goodwin & Duranti (1992, p.22) argue that 'the organization of human interaction is central to the analysis of context in a number of different ways' including features of face-to-face interaction that produces talk, the collaborative work of the participants in interaction that reflects the social organization, and so on. Duranti (1992, p.80) notes that the relation between words and the context of their use is a complex and dynamic one. Words do not reflect 'a taken-for-granted world 'out there'', but they help shape that world by defining the relationship between the speaker, the hearer, the referents, and the social activities.

To capture the uses and functions of lexical choices we must conceive context and language as a dynamic and evolving relation in which words mediate between different versions of the world and often let more than one version coexist in the act of speaking. (Duranti 1992, p.95)

For this reason, I consider that an accurate analysis of my subjects' lexical diversity in the classroom will depend on understanding the classroom context to be able to interpret differences with those produced by the same subjects on the IELTS speaking test. Understanding that the classroom should be regarded as a context in its own right, and not a replication of other contexts outside the school, is an essential pre-requisite to interpret classroom interaction and its effect on the linguistic choices made by the participants in this study. A detailed examination of some extracts of class transcripts while adopting an emic perspective that explores interaction from the participants' viewpoints will help to realize the interactional structure or 'interactional architecture' (Seedhouse, 2004) which is specific to the L2 classroom context in particular and the institutional context in general. It will help to conduct a qualitative analysis of the subjects' talk with a focus on their lexical performance as they monitor interaction, and as they organize turn-taking and use repair strategies in orientation to their pedagogical goals. Seedhouse (2004) calls his model of context a 'three way view', and he represents it in three concentric circles (Figure 2.4). The first view starts at the inner circle which is the 'micro context'. At this level, the focus is on the 'microinteraction' where the context is unique and the focus, at this stage, is on a heterogeneous view of interaction. The sequential organization is revealed through turn-by-turn examination, and the technical characterization of context is derived from the talk

details. Then, the pedagogical focus and the way interaction is organized in relation to it are specified. At the second level (or circle), the view broadens whereby a particular instance combining a pedagogical focus with an organization of interaction is compared to other similar instances in a specific L2 classroom context. At the third level of this model, context is viewed from an institutional perspective where the properties of an instance of interaction is compared to the properties of other instances of the L2 classroom interaction. At this level, Seedhouse purports that there is a degree of similarity due to three institutional properties that he identified, namely the dual role of language, the reflexive relationship between pedagogy and interaction, and the teacher's evaluation of students' performance. Those properties will be revisited in the next section (2.4.2) on the architecture of classroom interaction. The focus is on a homogeneous view of interaction. Seedhouse indicates that there is always a 'tension' between a homogeneous and a heterogeneous view of L2 classroom interaction.

By applying his model on a an extract from his database, Seedhouse demonstrates, at the micro context level, that it is unique in that the responses that the teacher gets from the learners are in all cases different and unpredictable. At the L2 classroom context level, Seedhouse explicates that it is similar in many ways to other instances in the database where repair, turn taking and sequence organization are conducted according to the pedagogical focus of the teacher. Finally, at the institutional level, the three interactional properties of institutional discourse are manifested in the fact that the teacher uses the second language for both managing interaction and correcting students' responses. It was, also, manifested in the fact the teacher rejected acceptable answers from students, because they did not meet the target forms of interaction which were planned as the pedagogical focus for that specific lesson. The third manifestation was in the repeated repair initiations made by the teacher, implicating negative evaluation of students' responses. Seedhouse adds that 'context is not seen as something external to the interaction or lurking in the background' (2004, p.213) but it 'inhabits the talk' as depicted by Schegloff (1993, p.114). Seedhouse contends that while his model demonstrates that all instances of interaction in the L2 classroom have the same properties and the same sequence organization, it shows their extreme diversity, fluidity and complexity. It also shows that the 'L2 classroom interaction is an institutional variety of interaction, and L2 classroom contexts are subvarieties' (2004, p.214). He believes that his three-way model can portray and analyse the three different levels simultaneously when each one feeds into the other in a reflexive way.

Walsh (2011, pp. 24-25) defines context as ‘both the background against which an event took place and the language used in that event’. Along with Seedhouse’s interactional architecture, Walsh’s features of classroom discourse will be used to examine the subjects’ talk to discern the impact of the classroom context on the lexical diversity of their speech. Walsh (2013) identifies six main features of classroom interaction which are, *teacher’s control of the communication, speech modification, elicitation, repair, student-student interaction, and computer-mediated communication*. However, for the purpose of this study which focuses specifically on teachers’ discourse and on the impact of features of interaction on their lexical diversity, the sub-sections following Seedhouse’s (2004) architecture of classroom interaction will review the first four features only, because the teacher’s role is central to them. Then, *scaffolding* and *repetition* will also be reviewed as two features of the L2 classroom interaction that characterize the teacher’s talk, and that can have direct influence on the lexical diversity of their speech.

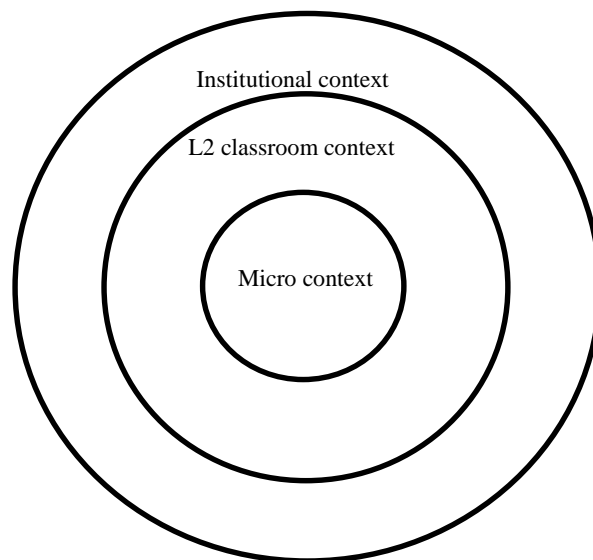


Figure 2.4 *A three-way view of context (Seedhouse 2004, p.210)*

Used with permission of John Wiley & Sons

2.4.2 The architecture of classroom interaction

In sketching the interactional architecture of the L2 classroom, Seedhouse (2004) indicates that the core institutional goal is teaching L2. He identifies three interactional properties that stem from the institutional goal and that shape interaction:

- 1- L2 has a dual role as the means and as the target of instruction.
- 2- There is a reflexive relationship between pedagogy and interaction, i.e. the interaction varies as the pedagogical focus changes.
- 3- Learners' linguistic forms and patterns are subject to the teacher's evaluation, though not necessarily in an overt way.

Seedhouse claims that those three properties are universal as they apply to all L2 classroom interaction, and they represent the 'foundation of the rational architecture and of the unique institutional fingerprint of the L2 classroom' (2004, p.187). He also posits that there is a basic sequence organization that applies to all L2 classroom interaction:

- 1- Introduction of a pedagogical focus
- 2- A minimum of two people are involved in interaction while orienting to the pedagogical focus.
- 3- Interactants get involved in analysing the pedagogical focus while performing turns in L2. Other participants produce more turns based on their analysis of the observed turns and their relation with the pedagogical focus.

Seedhouse (2004) identifies three actualizations of this sequence in his data base. First, based on the pedagogical focus announced by the teacher, the learners analyse the target production and try to produce it. Then, the teacher evaluates the appropriateness of that production in comparison with the announced pedagogical focus. In case of partial or total mismatch, the teacher conducts repair until the target linguistic patterns of interaction are produced. After analysing and evaluating, the teacher may introduce another pedagogical focus that the learners analyse and then try to produce new interaction patterns to meet it, and so on. Second, in pair or groups work activities, learners produce turns in L2 based on their analysis of the pedagogical focus introduced by the teacher. They listen to each other, they relate what they hear normatively to the pedagogical focus, they analyse it, and they respond to it. Teachers are physically absent, but remain present through the pedagogical focus that they have introduced. Third, learners may suggest their own pedagogical focus that can be accepted or rejected by teachers. If teachers accept, they have to analyse the pedagogical focus of the learner and take a turn to respond in normative orientation to that focus.

These three actualisations of sequence organization interact with features of interaction that are specific to the L2 classroom to give us a unique type of discourse. This uniqueness justifies why the main argument of the present study puts IELTS-generated discourse into question in terms of its representativeness of classroom talk. The following sub-sections will review six main features of classroom discourse.

2.4.3 Control of the interaction

The roles of teachers and students in class interaction are ‘asymmetrical’, according to Walsh (2011, 2013), since teachers have the upper hand in most class situations including those that are purportedly student-centred classes. Teachers are in command of class interaction patterns as they control both the topics of class conversations and turn-taking mechanisms. Usually, teachers decide on the time when students can speak, to whom they may address their speech, and how much time they can take to communicate their messages. Moreover, teachers can decide when to interrupt a student and take over a turn or give it to someone else; they direct and redirect discussions according to what they think is more appropriate for a lesson or for a class situation, and they change topics in accordance with their lesson plans. For all of these reasons, teachers talk more than students and they manipulate most of the class interaction. For Walsh (2011), teacher-dominated interaction may be summed up in a three-dimension classroom discourse similar to Sinclair & Coulthard’s (1975) IRF (Initiation-Response-Feedback) structure or to Mehan’s (1979) IRE (Initiation-Response-Evaluation) structure.

Myhill, Jones & Hopper (2006, p.14) report that teachers are frequently described as orchestrators of classroom interactions ‘conducting the responses from the class, signalling who should contribute, and controlling the outcomes’. That orchestration causes ‘considerable asymmetry’ in classroom talk due to the authority of teachers and their responsibility in making sure that learning takes place. Like Walsh (2011), they associate teacher dominance with the IRF exchange structure.

Heritage (1997, p.176) identifies this type of asymmetry in the contribution of teachers and learners in classroom interaction as *asymmetry of participation* which is part of institutional interaction where ‘there is a direct relationship between institutional roles and tasks, on the one hand, and discursive rights or obligations on the other’. Teachers, as institutional

representatives, ask questions and require students to answer them, which gives them the right to decide on three subsequent steps:

- 1- When to conclude a topic
- 2- What the next topic will be
- 3- How to shape the new topic (through questions)

Ellis (2008, p.797) attributes teachers' control of interaction to the prevalence of questioning in a classroom. Long and Sato (1983) found that ESL teachers asked far more display questions that test the learner than referential questions that seek information. Other studies like White & Lightbown's (1984) confirmed the same finding. The use of display questions restricts students' contributions in exchanges and keeps the teacher in total control of the interaction.

Seedhouse (2004, p.101) considers that the teacher's control of interaction is a factor of the reflexive relationship between the pedagogical focus and the organization of turn taking and sequence, for example in form-and-accuracy contexts the teacher exerts tight control over turn taking and sequence organization. The teacher controls students' turns and contributions, and may conduct repair if students produce forms other than the targeted ones, even when they are acceptable forms. In meaning-and-fluency contexts, students are given more interactional space to express meaning, and the organization of interaction is more flexible than in form-and-accuracy contexts.

2.4.4 Speech modification

Speech modification is one of the most discussed features in the literature on interaction in the L2 classroom, due to its fundamental role in making meaning more accessible for language learners (Hatch, 1978; Long, 1983 & 1996; Pica, 1994; Gass, 1997; Lightbown and Spada, 2006). Teachers try different ways to get their messages through, including simplification of language structures, choice of high-frequency vocabulary, using voice and intonation cues, body language, and so on to accommodate their speech to their students' level. At this point, teachers' simplification of their speech overlaps with Audience Design that was reviewed in section (2.2.4). Bell's (1984) Addressee Design comes into play to analyse and discuss the kind of simplifications that the subjects of the present study make in the classroom in order to accommodate their speech to their young learners. The focus will be placed on the impact of

simplification on the lexical choices of the subject and the resulting effect on their lexical diversity scores in the classroom. This will help to address the central argument of this research study which puts into question the predictive validity of IELTS scores to predict classroom performance for teacher trainees. Based on this fact, it is necessary to review speech modification as a salient feature of the L2 classroom.

According to Walsh (2013, p.31), one of the most important characteristics of classroom discourse is teachers' modification of their speech to ease the flow of communication, especially when students face difficulties responding to a teacher's initiation. McLaughlin (1985) indicates that in communication breakdown when teachers do not get immediate responses to a teacher-initiated interaction, they apply 'recovery or repair work aimed at recalibrating the interactional sequence' (McLaughlin 1985, p.150). A 'recovery' strategy aims at aligning the expected student answer through teacher prompts, repetition of questions, and even simplification of question wording to get the student to provide the appropriate answer. Simplification is performed through the use of basic grammatical structures and redundancy is reflected in the repetition of words, entire utterances or their components. Other features of speech modification include higher pitch of voice, exaggerated prosodic contours, baby-talk words, phonological simplification, and especially obvious abundance of interrogative forms. Long (1983, p.126) indicates that research on the speech of native speakers directed to non-native speakers found that native speakers use 'reduced or 'simplified' language in the form of 'shorter utterances, lower syntactic complexity, and avoidance of low frequency lexical items and idiomatic expressions'. In fact, in the context of the present study which is the L2 classroom in the UAE, avoidance of low frequency words and repeated use of more frequent and 'familiar' vocabulary to ease interaction in the classroom is regarded as a good practice and a sign of a teacher's success in adjusting their language to their students' levels. It is this teaching requirement that motivates the main argument of this study which posits that IELTS scores could be true indicators of student teachers' ability to use diverse vocabulary in academic contexts, however they may not be appropriate indicators of student teachers' ability to adjust that diversity to the level of their young addressees in a L2 classroom.

According to Lightbown and Spada (2006), pioneers of the interaction hypothesis including Hatch (1978), Long (1983, 1996), Pica (1994), and Gass (1997) believe that conversational

interaction is a basic condition of second language acquisition. The interactionist hypothesis focuses on the central role of speech modification by speakers to help learners understand and participate in conversations.

Walsh (2013, p. 31) postulates that the modification strategies that teachers use are not ‘accidental’ but ‘conscious and deliberate’. They happen for the following reasons:

- 1- to make sure that students understand what is communicated by the teacher in order to ensure effective learning.
- 2- to expose students to appropriate modelling of the target language.
- 3- to cater for the individual differences of the students and make sure that everyone is on track.

Along with the above-mentioned strategies, Walsh (2011, p.9) indicates that teachers use other strategies to clarify meaning, to check understanding and to confirm comprehension. Teachers use different techniques such as:

1. *confirmation checks* to make sure they understand correctly what the students mean
2. *comprehension checks* to verify that the students understand what is communicated to them by the teacher
3. *repetition*
4. *clarification requests* through which teachers ask students to clarify their speech or their ideas
5. *reformulation* of students’ speech by rephrasing their answers
6. *turn completion* by completing students’ utterance when they fail to do so
7. *backtracking* by recalling a previous part of a dialogue.

Chaudron (1988) surveys many studies that examined modification in teacher talk by comparing performances in different contexts and with different types of learners (native and non-native speakers). The comparisons covered different features including phonology, vocabulary, syntax and discourse, and while all of them are important to perceive how interaction and teacher talk look different in L2 classroom, only the ones that focus on vocabulary are going to be reported to keep focused on the central argument of this study which is the difference in lexical diversity between the classroom context and the test context. Though Chaudron (1988) questions the

reliability of some of those studies and though only few of them focused on vocabulary, it is important to signal out Mizon's (1981) study cited in Chaudron (1988) that compared teacher talk at the primary level for the same teacher and the same lesson with L1 and L2 learners. The results showed, among other findings, that there were lexical differences in that the teacher used less variety of content and function words and more proper nouns in L2 classroom than in L1 classroom. Another study was conducted by Henzl (1973) cited in Chaudron (1988) that compared teacher talk of eight teachers at university level for the same task with native and non-native speaker groups demonstrated that speech modification was marked by using more basic vocabulary that was stylistically neutral and less colloquial with L2 learners. Chaudron reports his own study (1982) in which he observed the speech of a native-speaking teacher on the same subject and on the same day for a native and a non-native class. He noticed that the teacher used more general high frequency vocabulary with non-native speakers, such as 'hold on very tightly' instead of 'clinging' (1988, p. 72).

Chaudron (1988, p.85) summarizes the findings of all previous studies on modification in the following seven points:

- 1- Rate of speech is slower
- 2- Pauses are more frequent and longer
- 3- Pronunciation is exaggerated and simplified
- 4- Vocabulary is more basic
- 5- Degree of subordination is lower
- 6- More declarative statements are used than questions
- 7- Teachers self-repeat more frequently

Speech modification is one of the features of interaction, but it is sometimes a technique that is used by the teachers to elicit information. Elicitation takes different forms as discussed in the following section.

2.4.5 Elicitation and questions

Elicitation is one of the most frequently used teaching techniques to collect information. Sinclair & Coulthard (1975) define elicitation as a classroom discourse act that requires a linguistic response. It takes an interrogative form, and it functions as a question. Though it sometimes

happens as a student turn to solicit the teacher's response, it is mostly associated with the teacher's turns. For this reason, it is important to review elicitation and examine its impact on the lexical diversity of the subjects in this study. In fact, elicitation could be a determining factor in keeping the subjects' lexical diversity scores at a lower level compared to their scores on IELTS due to its pedagogical role in eliciting responses from the students. In elicitation, teachers usually tend to keep their language accessible for the learners by using high frequency words, among other strategies. Mehan (1979) associates elicitation with the main instructional phase of the lesson. He identifies four different types of elicitation:

- a) *Choice elicitation*: students are asked to agree or disagree with a statement provided by the teacher.
- b) *Product elicitation*: students are expected to provide factual response such as a name, a place, etc.
- c) *Process elicitation*: students are expected to give their opinions or interpretations.
- d) *Metaprocess elicitation*: students are expected to reflect on the process of making connections between elicitations and their responses. (students are asked to provide the process by which they arrived at the right answer). Mehan notices that they are not frequent. (they represent only 1% in his study).

Walsh (2011, p. 33) claims that most classroom discourse is based on display questions, for which the answers are often known to both the students and the teachers. Such questions, make the classroom a 'unique social context' because most of the answers for those questions are already known in advance. He identifies five specific purposes for display questions including elicitation. In addition to display questions, Walsh (2013) identifies open-ended questions that teachers use to engage students in discussions requiring natural responses which are longer and use more complex forms. They are also known as referential questions. Chaudron (1988, p.127) points out that referential questions 'would promote greater learner productivity', and he reports that a number of research studies (Long and Sato,1983; Pica and Long, 1986; Dinsmore, 1985; Early, 1985; and Ramirez et al., 1986 cited in Chaudron, 1988) found that ESL teachers used more display than referential questions. He concludes that 'the more language-oriented the classroom, the more the teacher finds it appropriate to elicit linguistically constrained student contributions in order to promote practice in the language'.

Tsui (1992, p.101) identifies six subcategories of elicitation that are associated with specific functions in discourse:

- 1- *Elicit:inform*: it is meant to collect information from the addressee that is supposed to be unknown to the speaker until the answer is supplied.
- 2- *Elicit:confirm*: it is meant to seek the addressee's confirmation of the speaker's assumption.
- 3- *Elicit:agree*: it is meant to seek the addressee's agreement with the speaker's statement that involves some kind of commonly-shared assumption, or 'self-evidently true' assumption (Tsui 1992, p.107).
- 4- *Elicit:commit*: it is meant to obtain some kind of commitment on the part of the addressee for further interaction. For example:
A: Can I ask you a question?
B: Sure. (Tsui 1992, p.108)
- 5- *Elicit:repeat*: it is meant to ask the addressee to repeat a statement that they have just produced using words or expressions like *what? what did you say? sorry? huh?*, etc.
- 6- *Elicit:clarify*: it is meant to seek clarification for a statement that the addressee has just produced using words or expressions like *where? who? What do you mean?*, etc.

Weng (2009) cited in Kao, Carlin & Hsu (2011) finds that Tsui's (1992) categorization of elicitation missed two basic classroom-based sub-categories, because Tsui's work was based on social conversations between interactants of equal status. Weng adds two sub-categories to Tsui's list:

- 1- *Elicit:pseudo*: it is meant to check the addressee's (the student) understanding of some taught material. (equivalent to display questions mentioned above).
- 2- *Elicit:understanding check*: it is meant to check if the addressee is following the teacher's instructions.

Along with elicitation that generates students' responses and impacts teachers' lexical diversity scores, repair is another feature of the L2 classroom that pervades student-teacher interaction and can have a potential effects on the lexical diversity scores. Repair is discussed in the following sub-section.

2.4.6 Repair

The ways in which teachers react to students' errors are referred to as 'repair' strategies. Repair strategies or error correction strategies are essential for L2 classrooms. Teachers use them at different levels of frequency depending on the teaching and learning approaches they embrace, on the stage of the lesson, and on the level of students. English language teachers in UAE are no exception to this reality as they rely on repair strategies especially at the beginner and intermediate levels to help their students develop their communication skills and achieve better learning. In order to gain better understanding of the impact of repair strategies used by the subjects in this study on the lexical diversity of their talk, we need to investigate this feature of classroom interaction.

Sacks, Schegloff & Jefferson's (1974) identify four trajectories of error correction:

- 1- Self-initiated self-repair
- 2- Other-initiated self-repair
- 3- Self-initiated other-repair
- 4- Other-initiated other-repair

Schegloff, Jefferson & Sacks (1977) demonstrate that there is a clear preference for self-initiation in repair organization which is conducted by the learners themselves. Self-repair which is initiated within the same turn containing the problem is usually done in a successful way. However, most other-initiated repair that starts in the following turn takes more than just one turn to get accomplished, which might involve more teacher talk and more repetitions to get the learners or their classmates to fix the error.

Schegloff, Jefferson & Sacks (1977, p.376-377) find that other-initiation 'overwhelmingly' causes self-repair by locating the trouble source and giving the originator of the error another chance in the turn that follows to repair the problem. They conclude that the organization of repair in conversations favours self-correction, that can be achieved via well-organized self-initiation and other-initiation paths.

In his article on classroom talk, van Lier (1988) adopts this view of repair organization developed by Schegloff, Jefferson & Sacks (1977) and invites teachers to promote student self-

correction in their classes by increasing wait time and initiating repair instead of giving right answers. He focuses on the affective factor associated with repair strategies and its impact on the students' reaction to them. If the activity is focusing on a linguistic goal, a repair strategy that is focused on a linguistic form would normally be accepted positively, but if the activity is focused on a content goal a linguistic repair is most likely to produce a negative reaction and would, therefore, be resented.

As in control of interaction, Seedhouse (2004) suggests that there is a reflexive relationship between repair and the pedagogical focus. For instance, in form-and-accuracy contexts, repair is very much related to the linguistic forms that a teacher expects students to produce in order to meet the pedagogical focus of the lesson. Even when students produce answers that are linguistically and sequentially acceptable, teachers may initiate repair if they do not meet the targeted linguistic forms. The most common type of repair in this context is other-initiated self-repair, but other-initiated other-repair is common as well.

Seedhouse (2004) identifies eight ways of conducting mitigated negative repair in a form-and-accuracy context:

- 1- Using next-turn repair initiator using 'what?', 'sorry', 'pardon'
- 2- Repeating the word or phrase or part of word used immediately before the error.
- 3- Repeating the question or initiation
- 4- Repeating the student's wrong answer with a rising intonation
- 5- Supplying the right answer
- 6- Explaining why the answer is wrong without stating that it is wrong
- 7- Accepting the wrong form then giving the right form
- 8- Inviting other students to provide repair

What is noticeable here is that three of the eight ways use repeats in repair, which can be significant in reducing lexical diversity scores, especially that most of the classes in the present study focus on form and accuracy. In meaning-and-fluency contexts, Seedhouse (2004) notes that repair focuses on negotiating meaning and trying to reach mutual understanding. Correction is conducted in a way which is similar to ordinary conversation that van Lier (1988) calls 'conversational repair' as opposed to 'didactic repair' in form-and-accuracy context.

In his examination of an extract of classroom discourse with six pre-intermediate adult learners from Japan, Korea, Brazil, and Russia, Walsh (2011) notices that the teacher controls the use of language by matching pedagogic and linguistic goals. Errors are corrected on the spot and in a very direct and quick way. Walsh finds this way less time-consuming and less intrusive than the indirect way, though, like van Lier (1988), he does not recommend it in oral fluency work to maintain the flow of communication.

Whether it is applied in one way or another, the main role of repair remains to help learners in the different learning contexts to develop their understanding of how L2 works, in order to learn to use it appropriately. Repair in this sense works in harmony with scaffolding which aims to improve learning through ‘calibrated’ support. The next sub-section reviews scaffolding in the L2 context and provides background for the discussion of its effect on the lexical diversity scores of the student teachers in this research study.

2.4.7 Scaffolding

Though Walsh (2011, 2013) does not present scaffolding as one of the features of classroom interaction but as part of the sociocultural theory and its applications in the classroom, I am presenting it as a major feature of the L2 classroom discourse, especially with reference to the UAE context as a typical example of the L2 setting. English language teachers do a lot of scaffolding work to help their students cope with the linguistic demands of the tasks assigned to them. Therefore, reviewing scaffolding and examining how it can impact the lexical diversity scores of the subjects in the present study can represent a strong tool that helps to understand potential discrepancies between their oral performances in class and on the IELTS speaking test.

Walsh (2013, p.9) defines scaffolding as ‘the linguistic support given to a learner’, and with reference to Bruner (1990) he specifies that the support is given to the extent where the learner internalises learnt knowledge and makes conscious use of it. With reference to Doyle (1986), Walsh adds that scaffolding helps to reduce learning ambiguities. Challenge and support are two central concepts related to scaffolding, the purpose of which is to maintain interest and involvement for the first one, and to help understanding for the second one. Deciding on the amount of support to be given to the learner ‘requires great sensitivity and awareness on the part of the teacher’ (Walsh 2013, p.9). In communication breakdown when learners fail to find the

right word or to use the right strategy, the teacher intervenes to help learners sustain their interaction. Timing and sensitivity to students' needs are very important for effective scaffolding. Walsh compares teacher scaffolding with parents' help to their children when they struggle with word search and emphasizes that it requires active listening and minimal use of language on the part of the teacher. According to this view of Walsh, and in relation with the main argument of the present study, teachers are not supposed to display their lexical diversity but to choose the vocabulary that approximates the learners' level in order to ensure effective learning. In his SETT (Self-Evaluation of Teacher Talk), Walsh (2013, p.84) identifies three types of scaffolding: (1) '*reformulation*' that the teacher uses to rephrase a learner's contribution; (2) '*extension*' that is used to extend a learner's contribution; and (3) '*modelling*' to provide an example for the learner. In an analysis of a classroom extract, Walsh (2011) demonstrates that the teacher makes twelve turns, ten out of which contribute in engaging learners and prompting more complex and longer turns. He attributes this efficacy in scaffolding to the teacher's verbal behaviour that deals with students as active partners in the discourse. The teacher intervenes only when necessary to give support and to correct errors or to give her personal comment. This role of the teacher in scaffolding requires that the subjects in the present study use vocabulary that is congruent with their students' level or exceeds it at a minimal level to give them the necessary support for their progress in learning English language. Consequently, they will not use the range of vocabulary that they dispose of, and that they demonstrate on a proficiency speaking test, but they will restrict themselves to the vocabulary that is known to their young learners or that is prospected to be learnt.

Bruner (1983) gives a detailed explanation of the notion of scaffolding that he founded based on Vygotsky's Zone of Proximal Development (ZPD). He refers to Snow and Ferguson's work (1977) cited in Bruner (1983) that showed that parents' talk with their children does more than just modelling, but it involves refining their children's talk. 'Parents speak at the level where their children can comprehend them and move ahead with remarkable sensitivity to their child's progress' (1983, p.38). He suggests that input is 'arranged' by adults in the environment of the child to facilitate dual processing of concepts and communicative functions. This kind of support, Bruner calls 'adult Language Acquisition Support System' (LASS). He identifies four ways in which LASS works:

- 1- highlighting features of the world that the child is familiar with, and using related simple grammatical structures and typical linguistic forms to help the child relate the utterance with the mentioned event.
- 2- Modelling alternative lexical and phrasal forms for familiar gestures and sounds.
- 3- Using the ‘pretend’ situations that children act in play for language learning.
- 4- Using ‘conventionalized’ game formats to extend lexicon.

In scaffolding, teachers usually perform repeats of students’ utterances to reformulate or to model correct answers and statements in order for their students to improve their language skills. The following sub-section will review repetition as another important feature of classroom interaction that can potentially explain any discrepancies between the lexical diversity scores on the test and in the classroom.

2.4.8 Repetition

Repetition is another pervasive feature of the L2 classroom. Teachers use repeats of their own turns or their students’ turns for different reasons. In the context of the present study, teacher repeats can represent another important factor that explains why the lexical diversity in the classroom can be different than that displayed on the IELTS test. Fillmore cited in Ellis (2008) identifies repetition as an important feature of teacher talk that facilitates acquisition in L1 and L2 kindergarten classrooms. Park (2014) reports that previous research indicated that repetition is an important feature of L2 classroom interaction (Chaudron, 1988; Cook, 1994; Tomlin, 1994; Gass & al., 1998; Duff, 2000; Hellerman, 2003; Rydlaud and Aukrust, 2005; Yifat & Zardunaisky-Ehrlich, 2008; Piirainen-Marsh and Taino, 2009). Halliday and Hassan (1976) consider repeats as ‘question rejoinders’ which require a preceding statement or a command or elicitation of additional information. White and Lightbown (1984) found that teachers repeat or rephrase their questions as a kind of persistence to get answers. They found that ESL teachers ask up to four questions per minute, and that up to 64% of those questions are repetitions of previous questions. Seedhouse (2004) indicates that teachers use the strategy of repeating a learner’s wrong answer with a rising intonation to conduct repair without direct negative evaluation. Park (2014, p.147) indicates that repeats of students’ utterances can draw students’ attention to ‘key concepts, or linguistic forms, revoice a student’s contribution, correct it or affirm its validity and scaffold student learning’. The teacher’s repeats in the third turn usually

function as a request for an account of the student's response in the second turn, and often generate more student talk. According to Park (2014), a teacher's repeat in the third turn can be performed to confirm that an answer is correct and to reinforce for the whole class before moving on to another related sequence. He finds that third-turn repeats that promote further talk are features of a classroom that focuses on meaning-and-fluency. However, in form and accuracy contexts teacher repeats do not promote further talk as answers are usually fixed for display questions which are associated with this type of context due to differences in knowledge status between the teacher and the students. Park indicates that 'repeats make further talk from the student relevant but do not mobilize a response as strongly as a question would' (2014, p. 156)

Park (2014, p.159) distinguishes between repeats that ask for elaboration and repeats that confirm a response. Elaborative repeats occur when they follow 'unexpected or minimal responses to a previous question'. They can also occur when the previous response contains information that can be accounted for, or when the repeat is *ad hoc* rather than anticipated. Confirmatory repeats happen after correct responses, or when the question targets students' knowledge of language items, or else when it treats the previous response as correct on consideration. It can also occur when the answer is confirmed before moving to the next turn.

In this section on classroom interaction, I introduced the classroom context as a unique context with a unique architecture of interaction that lends itself to six important features, namely control of the interaction, modification of speech, elicitation and questions, repair, scaffolding, and repetition. Each feature was reviewed in the light of its potential impact on the lexical diversity of the subjects in this study and the subsequent significance of their IELTS speaking scores.

2.5 Language testing

This section on language testing theory and practice aims to situate the present research study within the discipline in order to discuss the main argument which calls into question the validity of IELTS test scores as predictors of performance in classroom situation. In order to achieve this aim, sub-section (2.5.1) briefly reviews the main milestones that marked the history of

language testing theory, sub-section (2.5.2) presents Bachman and Palmer's (1996, 2010) framework of language testing, and sub-section (2.5.3) discusses the development of IELTS test from its earliest version until the present day. The overview will provide necessary background knowledge about the test to discuss the validity of its scores for a teacher education program.

2.5.1. An overview of the development of language testing

In this section, I will shed light on the main developments in language testing theory that led to the current state of the discipline in relation with the developments in language proficiency and lexical proficiency theories that have been reviewed in the previous sections. The aim of the overview in this section is to show that the notion of testing has evolved historically, and that complex factors are involved in establishing or limiting the validity of language tests. A special reference will be made to the validity of IELTS scores as predictors of performance in school contexts. Davies's (1990) review of language testing history is used as a primary reference in this review as he draws clear and strong links between the history of language testing and IELTS development. In fact, the rationale for the format and the marking criteria of the speaking test as it is currently presented do not arise from a vacuum, but result from particular language and testing theories which have developed since the 1960s. Lado (1961) associates language testing with problematic areas that learners face in learning a language. The assumption is that the identification of problems facilitates language learning, therefore building a test on the identified problems is testing the language. Davies (1990) values the contribution of contrastive linguistics to the development of applied linguistics through language testing. In the 1960s language testing followed a structuralist approach that focused on language forms and accuracy. It was instrumental in the development of standardized English proficiency testing, like the English Proficiency Test Battery (EPTB) that Davies developed in 1965 and that represents the foundation of the current IELTS test. This will be covered in more detail in sub-section (2.5.3) on the development of IELTS.

The 1980s was marked by a clear shift towards communicative language testing that focused on integrative and pragmatic tests. Then, Bachman and Palmer (1996) presented their framework of language testing that became a reference for many practitioners in the field.

2.5.2 Bachman and Palmer's Framework of language testing

In their approach to language test development, Bachman and Palmer (1996) highlight the need for a correspondence between language test performance and language use:

If we want to use the scores from a language test to make inferences about individuals' language ability, and possibly to make various types of decisions, we must be able to demonstrate how performance on that language test is related to language use in specific situations other than the language test itself. (Bachman & Palmer 1996, p.10)

They argue that the necessity for a correspondence between language test performance and language use, justifies the need for a framework that identifies the main features of language test performance and of language use to be able to interpret performance on a language test as a special instance of language use.

In this respect, and in the case of the present research study, policy makers at HCT seem to have been under the assumption that there is a strong relationship between student teachers' language performance on IELTS and their performance in class. In other words, scores obtained on the test are considered as being reflective of the B.Ed. students' ability to perform successfully in the language teaching classroom. This assumption will be examined and discussed in light of the framework suggested by Bachman and Palmer (1996, 2010) to evaluate its appropriateness.

Bachman and Palmer's conceptual framework is based on the correspondence between two sets of characteristics (1996, p.12):

1) Characteristics of language use tasks and language test tasks

They indicate that it is necessary to consider the characteristics of a specific language use task, such as listening to an academic lecture in the target language, to be able to design a language test task that corresponds to it, and that can be used to make accurate inferences about a candidate's performance. Characteristics of a lecture can include the length of utterances, grammatical structures, textual features, functional and sociolinguistic factors, etc.

2) *Characteristics of language users and test takers*

Bachman and Palmer (2010, p.34) claim that it is necessary to know the characteristics of individuals to be able to identify the level at which those characteristics are engaged in language use tasks and language test tasks. They identify five main characteristics that interact with each other, which are *personal attributes, topical knowledge, affective schemata, cognitive strategies,* and most importantly *language ability*. Those characteristics of language users and test takers interact with characteristics of the language use setting or the test task which include among other characteristics the language, the physical setting, and other language users (Bachman & Palmer 2010, p.34). They also claim that those same characteristics need to be engaged in an assessment task to be able to generalize interpretations to other situations of language use.

Bachman (1990) and Bachman and Palmer (1996, 2010) contributed significantly to the development of communicative language testing theory. Their approach to language testing provides a sound theoretical background on which the design or the choice of a specific language test can be based. It represents a frame of reference for the present research study as it discusses the relationship between the test task and the language use task, which corresponds to the relationship between the IELTS speaking test task and the classroom teaching task. It also discusses the characteristics of the test task and the language use task, which will provide a theoretical background to examine the characteristics of IELTS speaking tasks and to compare them with the characteristics of the classroom teaching tasks. Bachman and Palmer's approach rests on two basic principles:

1- *the necessity for correspondence between language test performance and language use* to be able to make meaningful inferences of a test score about an individual's language ability.

If we want to use the scores from a language test to make inferences about individuals' language ability, and possibly to make various types of decisions, we must be able to demonstrate how performance on that language test is related to language use in specific situations other than the language test itself. (Bachman & Palmer 1996, p. 10)

To clarify that relationship, Bachman and Palmer (1996, 2010) establish a conceptual framework that defines the characteristics of language test tasks and language use tasks. They define language use as the creation or interpretation of meaning in individual discourse, or the

negotiation of meaning between two or more individuals. It includes two types of interactions: (1) interaction among the attributes of individuals, and (2) interaction between the language user and the characteristics of the setting (Bachman & Palmer 2010, p.34) . They define a language use task as ‘an activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation’ (1996, p.44). A test task, on the other hand, is an activity that allows test designers or users to make inferences ‘that generalize to those specific domains in which test takers are likely to need to use language’ (*ibid*). The task characteristic framework that Bachman and Palmer (1996) present for the description of language use tasks and language test tasks consists of characteristics of:

- a) the setting (physical characteristics, participants, time of the task)
- b) the test rubric (instructions, structure, time allotment, scoring method)
- c) the input (format, language of input)
- d) the expected response (format, language of expected response)
- e) the relationship between the input and response (reactivity, scope of relationship, directness of relationship).

The illustration of their model, clarifies Bachman and Palmer’s view of the correspondence between language use and language test performance. In the case of the present study, it is the strength or weakness of that correspondence between performance on IELTS as a language test task and performance in the classroom as language use task that drives part of the discussion in this research study.

2- *a clear definition of the qualities of test usefulness*, which include reliability, construct validity, authenticity, interactiveness, impact, and practicality, while considering their complementarity rather than their tension. Bachman & Palmer advocate that test usefulness cannot be evaluated in general terms but only with reference to a specific test and a specific testing situation. Bachman & Palmer argue that

any given language test must be developed with a specific purpose, a particular group of test takers and a specific language use domain (i.e. situation or context in which the test taker will be using the language outside of the test itself) in mind’ (1996, p.18).

They add that validity and reliability are critical qualities because they can justify the use of test scores to make inferences about an individual’s language ability or to make decisions.

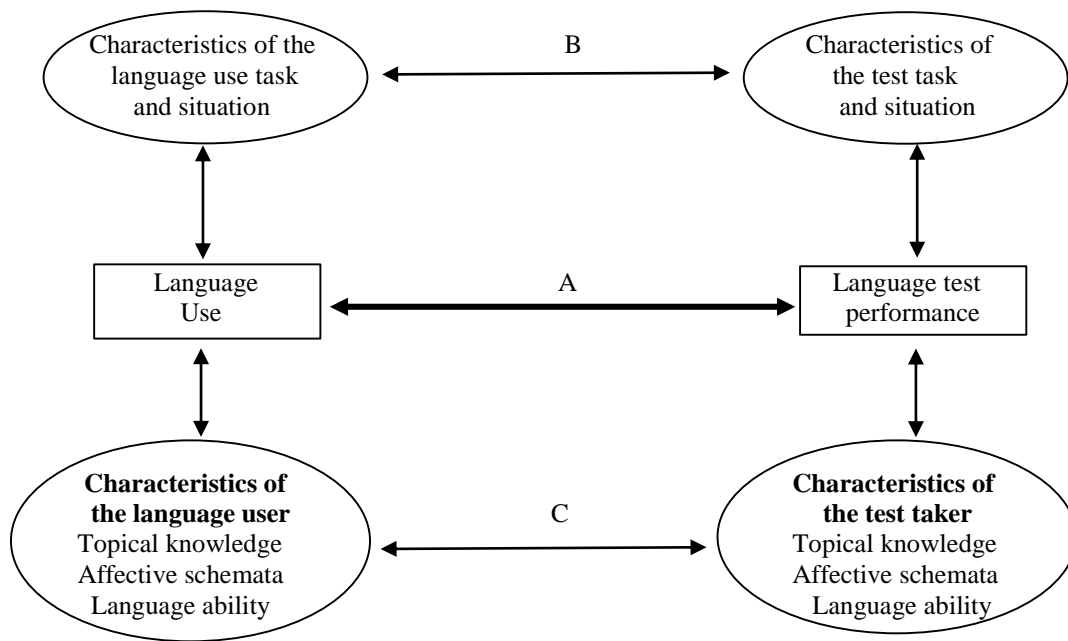


Figure 2.5 *Correspondences between language use and language test performance* (Bachman & Palmer 1996, p.12). Used with permission of Oxford University Press.

2.5.3 Testing speaking

In defining speaking as a construct, Fulcher (2003, p.18) links it with observable ‘things’ that can be scored and that can be operationally tested. In line with the hypothesis of this thesis which suggests that testing student teachers’ speaking ability should correspond with the context where they perform speaking which is the classroom, Fulcher (2003,p.19) indicates that recent developments in language testing call for considering contextual factors when they clearly impact the performance of the construct and related scores. With reference to Bachman and Palmer (1996), he postulates that a full description of the target language use domain (TLU) should be considered in designing a speaking test, because the inferences that are made of the scores obtained are not about the speaking ability in general but related to a specific context. Luoma (2004, p.30) agrees that context is central to language use, because it ‘includes concrete aspects of the situation such as the place where the talk happens, and cognitive and experiential aspects such as the language use experiences.’

For Fulcher (2003), it is the test purpose that should guide the definition, the range, and the generalizability of the construct of speaking, that is why ‘we should not assume that any description, any rating scale, captures some psychological reality that exists in the language competence of all speakers for all time in all contexts’ (Fulcher 2003, p.19). In fact, decision makers in the context of the current study seem to have been assuming that the IELTS speaking test can capture the reality of teacher talk in the classroom, but this research puts that assumption into question, a position that is supported by the findings reported in the following chapters. As stated by Fulcher (*Ibid*) different constructs are used for different purposes, therefore ‘the construct definition we decide upon should not be evaluated by its correspondence to psychological reality, but by its utility in making inferences from test scores’ (Fulcher 2003, p.20)

Fulcher (2003, p.24) claims that speaking has some specific features that make it different from writing. Some of those features include the opening and closing of conversations, adjacency pairs, and turn-taking. Interaction in speaking takes different forms that are mostly typical of speech. Rules of speech are related to the context in which it is produced and are affected by the formality of the context, the social status of the different participants, and other factors. Moreover, speaker do not have ample time to plan, but they have to produce speech that is appropriate for the situation. This depends on the automaticity of speech which comprises different factors including control of the structure of the language, the lexical range, the use of formulaic expressions, and ability to monitor speech effect on the listener. Automaticity is very much linked with the complexity of the message, the familiarity with the topic, the expected pace of processing, the required accuracy, and possible reactions to a wrong performance. According to Fulcher (2003, p.47) one of the most important challenges of designing speaking tests is *construct underrepresentation* which is a failure ‘to capture important aspects of the construct the test is intended to measure’. A second challenge is *construct-irrelevant variance* which is caused by factors that are not related to the construct that is meant to be measured. In both cases, the result is that inferences drawn from the test scores do not serve the main purpose of the test. What this research study investigates is the strength of relation between the purpose set up by IELTS developers for the speaking component of the test and the purpose conceived by the test users in the context of present study.

2.5.4 The IELTS test

The International English Language Testing System (IELTS) is a language proficiency test jointly created and managed by Cambridge ESOL examinations, British Council and IDP: IELTS Australia. It assesses four language skills, namely listening, speaking, reading and writing and it is available in two formats, academic and general. The test identifies nine levels of performance called bands and assigns them scores as follows:

9: *Expert user*

8: *Very good user*

7: *Good user*

6: *Competent user*

5: *Modest user*

4: *Limited user*

3: *Extremely limited user*

2: *Intermittent user*

1: *Non user*

0: *Did not attempt the test*

Each band level has detailed performance descriptors for the speaking and the writing components of the test. For the listening and the reading components, a band-score conversion table converts scores out of 40 into corresponding band levels.

The *listening test* lasts 30 minutes. Candidates listen to 4 recordings of native English Speakers:

- 1- Recording #1: A conversation between two people in a social context
- 2- Recording #2: A monologue in a social context
- 3- Recording #3: A conversation between up to 4 people in an academic or training context
- 4- Recording #4: A monologue in an academic context

The *reading test* lasts 60 minutes. Candidates read 3 'long texts which range from the descriptive and factual to the discursive and analytical' (IELTS, 2016). Different sub-skills are tested like reading for gist, reading for the main ideas, skimming, and so on.

The *writing test* lasts 60 minutes. Candidates are presented with two tasks: (1) write in response to an instruction about a given graph, table, chart, or diagram, and (2) write an essay in response to a point of view, a problem, or an argument (IELTS, 2016).

The *speaking test* will be described in more details in the following section (2.5.4-1) as it is an important component of the focus of this research study.

The markers are certified examiners who are monitored regularly (up to four times a year) by examiner trainers. They receive written feedback on their rating, and they may be called to review IELTS procedures if issues with their ratings are raised. Standardisation sessions take place every two years to recertify examiners. (IELTS, 2016)

Based on HCT graduation requirements, the B. Ed. students must achieve a *competent user* level which is equivalent to an overall band 6 score to be qualified as teachers of English in the UAE. A *competent user* is defined as a language user who ‘has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings in some situations. Can use and understand fairly complex language, particularly in familiar situations’ (IELTS Handbook 2007, p.4)

1) The IELTS speaking test

The speaking component of the IELTS test consists of three main parts:

- 1- Part 1 is *the introduction and interview*. In this part the examiner introduces himself/ herself and confirms the candidate’s identity then asks general questions about the candidates ‘themselves, their homes/families, their jobs/ studies, their interests, and a range of familiar topic areas’ (Official IELTS Practice Materials 2007, p.50). The duration of this part should be between 4 and 5 minutes.
- 2- Part 2 is *the individual long turn*. In this part the candidate is given a prompt card with a specific topic to talk about for 1 to 2 minutes. Before talking the candidate is given 1 minute to prepare his/her talk using prompts given on the card. By the end of this long turn, the examiner can ask one or two rounding-off questions. The duration of this part is between 3 and 4 minutes including the 1-minute preparation time.

3- Part 3 is the *two-way discussion* in which the examiner engages the candidate into a more challenging discussion that elicits more abstract thoughts but which is linked to the topic talked about in part 2. The time allotted to this part is between 4 and 5 minutes.

The overall time for the test is 11 to 14 minutes and the examiner is not allowed to reword the questions or give clarifications or verbal/non-verbal feedback in part 1 and 2 of the test. However, in part 3, the examiner is given more leeway to reformulate the questions according to the candidate's level and answers.

The candidate's performance is assessed according to IELTS speaking descriptors developed at a nine-band performance scale. The descriptors are divided into four main areas:

- 1- *Fluency and coherence*: this area focuses on the continuity of speech and the rate at which it is delivered by the candidate. It also assesses the logical organization of ideas and the use of cohesive devices. (IELTS Handbook 2007)
- 2- *Lexical resources*: this area focuses on the range of vocabulary that is produced by the candidate and on the ability to deliver meaning in a precise way. It also assesses 'the variety of words used and the ability to circumlocute (get around a vocabulary gap by using other words) with or without noticeable hesitation.' (IELTS Handbook 2007, p.12).
- 3- *Grammatical Range and Accuracy*: this area focuses on the candidate's ability to produce accurate and appropriate grammatical structures and communicative functions. It also assesses the length and complexity of produced forms and the ability to manipulate language to manage information focus. (IELTS Handbook 2007)
- 4- *Pronunciation*: This area assesses the candidate's ability to make his/her speech intelligible enough and the extent to which L1 influence affects the production of a comprehensible speech. (IELTS Handbook 2007)

2) *The development of the IELTS test*

This section overviews the evolution of the IELTS test to highlight the changes that have happened since its first inception until nowadays. The changes demonstrate that the test writers have always tried to keep the test abreast of all developments in second language learning and

testing, and have conducted validation revisions to ensure that the test achieves optimal levels of validity.

Though Taylor (2003) claims that Cambridge testing started as early as 1913 with CPE (Certificate of Proficiency in English), Davies (2008) indicates that Perren's (1963) battery of tests, which was used for both foreign and native speakers of English, played a seminal role in the development of academic English language proficiency testing in the UK. It was followed by Davies's (1965) English Proficiency Test Battery (EPTB), then Criper and Davies's (1980) English Language Testing Service (ELTS), and finally the International English Language Testing System (IELTS) in 1989. IELTS was established after the submission of Criper and Davies's (1988) ELTS validation report that emphasized two main aspects of the new test: (1) compromising practicality, and (2) maximizing predictive validity (Davies 2008, p.91). Since then, IELTS has gained growing international reputation as a useful language proficiency standardized test, especially for admission of overseas students to the British and Australian tertiary level institutions. However, IELTS has undergone several changes since its first administration in 1989. The first significant changes were suggested by UCLES (University of Cambridge Local Examinations Syndicate) revision specifications in 1993. The suggested changes covered:

- 1- *Management*: taking full control of the development of the test by UCLES
- 2- *Technology*: integrating new technology in developing computerized versions of the test
- 3- *Theory*: considering development in measurement theories with special reference to Messick's (1989) consequential validity, Bachman's (1990) framework of language testing and language proficiency, and Canale and Swain's (1980) communicative competence.
- 4- *Marketing*: improving marketing strategies.
- 5- *Administration*: tightening procedures to improve reliability.
- 6- *Validation and research*: developing procedures to capture and store data in order to improve test construction and 'post hoc validation' (Davies 2008, p.93).

The changes were officially operational in 1995, but subsequent and regular revisions of the speaking and writing components followed to keep the test up to date. The result was 'an

astonishing growth in the take-up of IELTS' (Davies 2008, p.99). The *IELTS Annual Review* was first published in 1995 with the aim to give updates on the test and its development.

Taylor (2001a, p.9) indicates that the 1995 changes did not include changes in the speaking module because the revision process took longer. The project was launched only in 1998 and the changes became operational only in 2001. The revision included the assessment criteria and the rating scale. The reviewers deconstructed the previously used holistic scale into four main subscales:

- 1- Pronunciation
- 2- Fluency and coherence
- 3- Grammatical range and accuracy
- 4- Lexical resource

The task design changed from having five phases to the current three-phase format

- 1- Introduction and interview
- 2- Individual long turn
- 3- Two-way discussion

The introduction of the Examiner Frame represented a 'significant change' (Taylor 2001b, p.3). It is a script for the examiner to follow in the interview with clear control especially in part 1 and 2 to make sure candidates receive the same input. However, in part three the examiner has more freedom to accommodate language to the level of the candidate.

3) *The issue of language sampling in the IELTS test*

Davies (2008, p.105) evokes the issue of sampling in language tests in general and in IELTS in particular. He indicates that it is 'inescapable' but at the same time it is problematic, because the chosen samples should match the language use that the candidates are supposed to encounter. In fact, this assumption of Davies legitimizes the main argument raised by this research study which seeks to investigate the match between the level of lexical diversity solicited by the IELTS speaking test and that of a classroom situation when the subjects teach young Emirati learners. This issue leads to the discussion of the argument-based approach (Kane, 1992) that

will be presented later in this chapter (section 2.6.5) in relation with the interpretation of IELTS scores.

Proficiency test sampling followed the development of learning and teaching approaches. First, the EPTB test adopted a structuralist approach which focused on sampling learners' difficulties in grammar and vocabulary. Then, ELTS followed the strong version of the communicative approach (communicative competence) focusing on real life examples that reflect language uses. Finally, IELTS adopted a 'moderate' approach that compromised between the first two approaches with a focus on Bachman's interactional ability. 'IELTS represents a kind of regression to the mean, a (good) compromise between the extremes of the structural and the communicative' (Davies 2008, pp.108-109). Davies considers IELTS a valid test of academic proficiency, because it uses texts that have features of academic language for the listening and reading skills, and it requires candidates to produce 'cogent and coherent discourse' for the speaking and writing skills (2008, p.113). He defines academic proficiency as:

Skilled literacy and the ability to move easily across skills. In other words, it is the literacy of the educated, based on the construct of there being a general language factor relevant to all those entering higher education, whatever specialist subject(s) they will study. (Davies 2008, p.113)

Despite this clear evidence that the test has been revised regularly to ensure that it is a valid test of academic proficiency, the question that remains to ask and that this research study tries to answer is: how much can scores obtained on such a 'valid' test be predictive of a candidate's performance in a specific context? The specific context we are concerned with in this research study is the classroom context.

2.6 Validity

In her review of historical perspectives related to validity, Chapelle (2012, p.23) focuses on three main references that contributed significantly to the development of the concept. First, she indicates that Lado's (1961) contribution to validation procedures remain important despite his out-dated views of language. She attributes this to the fact that he combined his own ideas with

both the contrastive analysis perspective and published work in educational measurement. Second, Messick's (1989) seminal work on validity that challenged the commonly known and simplistic definition of validity (whether the test measures what it is supposed to measure) had clear impact on research in testing. He developed four basic concepts that make up validity:

- 1- It is not a property of a test but the interpretation and use of test results
- 2- It is one unitary conception with construct validity as a central concept
- 3- It includes relevance and utility, value implications, and social consequences
- 4- Validation is an ongoing process of inquiry

Third, Bachman (1990) and Bachman and Palmer (1996) contributed significantly to the definition of 'construct' through their framework of language ability. They also introduced a new 'term and manner of conceiving validity' (Chapelle 2012, p.24) which is 'test usefulness'. It is the evaluation of the test based on the use for which it is initially designed.

While considering Messick's unitary definition of validity for the discussion of IELTS scores as a graduation requirement for the B.Ed. student, this section will start with a simplistic approach that discusses basic types of validity (2.6.1) and recent types of validity (2.6.2). The importance of these two categories of validity for this study lies in providing more specific arguments for or against the overall predictive validity of IELTS scores, especially with reference to the relationship between the context of IELTS test and the context of the classroom and to Cummins's (2000) framework of language proficiency that was reviewed in sub-section (2.2.2). *Predictive validity* as the main focus of investigation in the present study will be discussed at length in section (2.6.3). Messick's (1989) definition of validity will be discussed in more details in section (2.6.4) as it represents a strong frame of reference for most researchers in language test validity and for the present thesis as well. Finally, I will devise section (2.6.5) for Kane's (1992) interpretive argument that will be discussed as a potential solution for the validation of IELTS scores for a teacher education program.

2.6.1 Basic types of validity

The basic types of validity that were discussed in the 1950s and 1960s when language testing started to gain momentum in the field of applied linguistics include criterion-related validity,

content validity, construct validity, and most importantly -for the present study- predictive validity.

1) *Criterion-related validity*

Shepard (1993, p.410) notes that criterion-related validity is important for selection and placement decisions, because it evaluates ‘the correspondence between test performance and expected criterion performance’. However, the empirical relationship between the test performance and the criterion, though essential, is not enough to validate the test use, as some criteria could be invalid. Shepard calls for the evaluation of the relevance and the ‘integrity’ of the criterion measures, and their claimed predictive validity. In fact, the relevance of IELTS speaking test to the context of the B.Ed. program in the UAE will be evaluated in this research study.

2) *Content validity*

Content validity is usually associated with the history of achievement testing. It focuses on the knowledge and skills that students are expected to demonstrate, rather than the table of contents of a course (Shepard 1993, p.411). In this research study correspondence between knowledge of English language and skills demonstrated on IELTS and knowledge and skills of teaching English demonstrated in a classroom will be examined.

3) *Construct validity*

According to Shepard (1993), construct validity was first introduced in 1954 by the American Psychological Association (APA) who relate construct validation to theory-based predictions and empirical data that confirm or disconfirm predictions. Cronbach and Meehl (1955, p.282) define it as ‘the measure of some attribute or quality which is not “operationally defined”’. It cannot be identified by specific procedures but by the orientation of the researcher who decides on the construct reflected by the research instrument and its related meaning (Cronbach & Meehl 1955, p.290). Shepard advocates that construct validity was established through the application of the scientific method in testing to validate the interpretation of test scores. It relies on both the rational argument and the empirical verification (1993, p.416).

4) Predictive validity

As the main focus of this research study, predictive validity will be reviewed separately in subsection (2.6.3).

2.6.2 Recent types of validity

Davies et al. (1999) identify the four types of validity mentioned previously as the common types of validity, but they also identify other types like *face* validity, *consequential* validity, *systemic* validity, *discriminant* validity, *divergent* validity, and *ethicality* validity. However, to keep within the scope of this research study, I will not elaborate on all of them but I will highlight two recent types of validity that have a direct connection with the focus of the actual research study, namely, *cognitive validity* and *context validity* which are based on Weir's (2005) theory-based validity. The relevance of these two types of validity lies in providing a reference to a more specific comparison between the impact of cognitive demands on the performance of the subjects on IELTS and their performance in a classroom situation. They will also provide more arguments based on the correspondence between the contextual factors of each setting.

1) Cognitive validity

Drawing on Weir's (2005) work on theory-based validity which has recently been known as cognitive validity, Field (2011, p.65) defines the cognitive validity of a speaking test as 'the extent to which the tasks in question succeed in eliciting from candidates a set of processes which resemble those employed in a real-world speaking event'. It is also the extent to which those processes are graded according to the level of cognitive demands required from a candidate. In this respect, the IELTS speaking test in the current study is going to be examined and the impact of the cognitive demands on the subjects' choice of vocabulary items is going to be compared with the impact of the cognitive demands of the teaching situation. Field advocates that the validation of tests that measure performance necessitates the establishment of validity of the test task, test content, and test conditions that require the replication of the same cognitive processes happening in a natural context. This type of validity is important for tests that measure the general competence in any of the proficiency test skills, because the test value is related to its ability to predict a candidate's performance in an L2 context.

Thus it becomes important for test producers to know, and to be able to demonstrate, how far what happens in the testing situation replicates cognitive processing in the world beyond the test, so that test users can have confidence in the meaningfulness and usefulness of the score outcomes from the test. (Field 2011, p. 66)

Based on Levelt's (1989 & 1999) models of cognitive processing in speaking activities in an L1 context, Field (2011) suggests an adapted model that applies to the L2 context. The model considers the following factors that affect L2 speakers' performance on speaking tests:

- a- *Conceptualization* which includes macro-planning and micro-planning.
- b- *Grammatical-encoding* which includes the construction of a 'surface structure' and the conversion of the structure and the related lexis to phonological forms.
- c- *Phonological encoding* which gives planned speech a concrete form.
- d- *Phonetic encoding, articulation* which include a set of 'phonological representations in the mind' and a set of 'highly automatic processes attuned to the articulatory settings' (Field 2011, p.112).
- e- *Self-monitoring* which compares the impact of the speech with the set goals at the conceptualization stage.

2) Context validity

In connection with Field's (2011) definition of cognitive validity and based on Weir's (2005) socio-cognitive test validation framework, Galaczi & Ffrench (2011, p.112) investigate the context validity of Cambridge ESOL General English Speaking tests which include KET, PET, FCE, CAE, & CPE. They investigate the relationship between 'contextual parameters' of test tasks and proficiency levels, and they claim that it might be applied to other Cambridge speaking tests. They discuss the relationship between the test task with its contextual parameters and the performance of the test taker. They use Weir's (2005) test validation framework that provides the contextual parameters of a task in order to analyse the task features in a socio-cognitive approach. Weir's (2005) aspects of context validity for speaking include the setting which comprises the task itself (response format, purpose, weighting, criteria, order of items, and time) and the way in which it is administered (physical conditions, uniformity of administration, and security). It also includes the demands which comprise linguistic demands (channel of communication, discourse mode, length, nature of information, topic familiarity, lexical

resources, structural resources, and functional resources) and interlocutor demands (speech rate, variety of accent, acquaintanceship, number, and gender).

2.6.3 Predictive validity

Predictive validity concerns the extent to which inferences made of a test predict subsequent performance on a course or a particular job. It is usually associated with proficiency tests like IELTS and TOEFL. Candidates usually take the test in their home country, and then as they arrive and settle in the host country they are given a similar test, and a correlation between the results on both tests would reveal the degree of predictive validity. However, this type of validation has always come under criticism for what is known as the problem of ‘truncated sample’ due to the fact that candidates who fail the test in their home country do not sit for the equivalent test in the host country, which reduces ‘the spread of students’ scores’ and affects negatively the validity coefficient (Alderson et al., 1995).

Another inherent problem of predictive validity is that it overlooks the abilities being measured because it is difficult ‘to identify and measure all the abilities and factors that are relevant to the criterion’ and to be clear about the type and strength of the relation between the predictors (Bachman 1990, p. 252).

Research studies on the predictive validity of language proficiency tests have been inconclusive regarding their ability to predict academic achievement. The two following sub-sections will summarise some studies that reported positive findings confirming the predictive validity of IELTS, and some others that reported negative findings indicating its weakness.

1) *Studies reporting weak predictive validity of the IELTS test*

The following three research studies are examples of studies that reported that IELTS predictive validity is weak:

Cotton and Conrow (1998) investigated the predictive validity of IELTS by examining the relationship between IELTS scores and academic achievement of a group of international students at the University of Tasmania, Australia. They ran correlations between IELTS scores and students’ first and second semester results using three measures of academic performance, namely Grade Point Average (GPA), academic staff ratings of students’ performance (including

course tutors, student advisors, and English support tutors), and students' assessment of their own academic performance in the first semester and the second semester. Correlations between the three measures and IELTS overall scores revealed that there was no significant correlation, but the reading component of IELTS had moderate positive correlation with academic scores (GPA). Based on that, Cotton and Conrow (1998, p.97) conclude that the reading subtest of IELTS 'has the greatest ability to predict future academic performance', whereas other subtests had very weak or no link with academic performance.

In a similar study, Kerstjens & Nery (2000) conducted an investigation into the relationship between IELTS scores and subsequent academic performance of 113 first-year international students enrolled in the faculty of Business at an Australian university. The population was composed of two groups that were almost equal in size, one took Higher Education courses and the second one took Training and Further Education (TAFE) courses. Kerstjens & Nery (2000) found significant correlation between the reading and writing scores and the students' GPAs for the total population. However, when they examined the correlation for each group separately, they found that the scores correlated only for the reading skill for the Higher Education students and did not correlate for any skill for the TAFE group, though the writing skill for the latter group correlated higher than the Higher Education group. Kerstjens and Nery (2000) concluded that the predictive ability of the IELTS test in relation with students' GPA ranges from small to medium for the total population and the Higher Education group, but not for the TAFE group. The reading component of the test was the only significant predictor of academic performance for both groups.

In another research study that resembles to the present one in that it focuses on a teacher education program, Elder (1993, p.72) raises the issue of determining the required language proficiency level to succeed in a teacher education program that includes 'both academic and a school-based teaching component'. She investigated the relationship between international students' language proficiency as measured by IELTS and their performance in teacher education courses at different tertiary institutions in Melbourne. The research questions covered the following:

- (1) Whether IELTS is a reliable predictor of success in education courses.

- (2) The extent to which IELTS is accurate in comparison with each institution's screening procedures.
- (3) The highest IELTS 'threshold' to enter teacher education.
- (4) Whether scores obtained for the component skills of IELTS predict difficulties encountered by students in coursework tasks.
- (5) Whether exposure to and use of the second language during the practicum year have an effect on the relationship between predictions and outcomes.

While admitting that there is a lack of agreement in research about the ability of screening tests to predict subsequent performance, Elder joins Ferguson & White (1993) and Bellingham (1993) (cited in Cotton & Conrow, 1998) in asserting that there is some evidence that low levels of language proficiency can be stronger indicators of subsequent performance. Correlations of IELTS overall and component-skill scores at admission with two course ratings revealed that IELTS is a 'reasonably good predictor of short term performance' (Elder 1993, p. 78) as it correlated significantly –but not strongly- with first semester ratings. However, test scores correlation with second semester course ratings was insignificant which made Elder conclude that IELTS predictive power diminishes with time, and she attributes this to (a) improvements in language ability after a longer period of exposure to language and to instruction, and to (b) increasing role of non-language variables such as subject knowledge, interpersonal skills, cultural competence, and so on. Moreover, Elder finds that listening scores predict better first-semester performance than other skills. She attributes this finding to the probability that education courses place more emphasis on listening skills which are necessary both for academic lectures and classroom interaction. She makes a general conclusion that language proficiency seems to be a weak predictor of success in a teacher education program. She also, agrees with previous studies that IELTS can predict subsequent performance only at low levels of proficiency as she found that 'the strongest level of agreement between test predictions and academic outcomes occurred at the Band 4.5 levels' (Elder 1993, p. 87).

2) Studies reporting strong predictive validity of the IELTS test

Ingram and Bayliss (2007, p.4) investigated the language behaviour of two groups of international students who came from a non-English-speaking background in their first semester at an Australian university. The investigation aimed at comparing their language behaviour in different tertiary contexts with their IELTS entry scores to see if there was any matching, and if their proficiency level, as determined by their IELTS scores, was enough to cope with the academic language tasks in their studies. Their population included 28 international subjects who were in their first semester of studies at two Australian tertiary institutions. The findings revealed that 25 out of the 28 participants exhibited language behaviour that equalled or exceeded what IELTS scores predicted. Ingram and Bayliss (2007, p.60) concluded that IELTS scores ‘can quite accurately’ predict students’ language behaviour in the first semester of their studies.

Paul’s (2007) research study complemented the previous study of Ingram and Bayliss (2007). It focused on four participants from the population of the previous study, and it compared their language behaviour in speaking and writing -as indicated by their IELTS scores- with their first semester language performance in different university programs. Paul (2007, p.4) used discourse analysis of classroom and IELTS spoken and written tasks in order to examine ‘how aspects of language such as complexity, accuracy and fluency in academic settings change from that produced under IELTS test conditions’. The findings of the four case studies revealed that language production was generally similar or improved in writing for three of the subjects, and it was generally similar as well in speaking though two of the subjects showed lower level in some aspects of academic tasks. The overall conclusion was that the subjects demonstrated similar level of language in classroom tasks as that exhibited in IELTS test, and therefore IELTS scores generally predicted students’ language levels in academic contexts.

This third case study does not investigate the predictive validity of IELTS, but a similar language proficiency test that is designed in the same context as the present study –the United Arab Emirates. Rumsey (2013) conducted a research on the predictive validity of the Common Educational Proficiency Assessment (CEPA) that is used as a fundamental requirement for admission to one of the three government tertiary institutions in UAE. The test scores are used to place candidates in the appropriate level of the pre-bachelor program called ‘foundation’

which lasts one year and is meant to prepare students with limited English, Maths and computer skills to cope with the demands and challenges of the diploma and bachelor programs. Rumsey (2013) conducted her study at the Higher Colleges of Technology which is the same institution where the actual research study has been conducted. She indicated that prior to her study, CEPA had been reported to have had relatively high correlation with students' English marks at the end of the one-year foundation. Rumsey's aim of her study was to examine the predictive validity of CEPA and to identify other variables that contributed to having a fairly strong predictive validity. She used a mixed method combining qualitative and quantitative data-gathering tools from 347 students, foundation supervisors and coordinators, the central administrator and the CEPA supervisor. In her analysis of the data, she found that CEPA was a good indicator of students' level in English prior to their admission to the college, and it was a good indicator of students' subsequent performance at the foundation year.

A third category of research studies investigated the use of IELTS scores as cut-off scores for admission to higher education institutions. The relevance of those studies to the present one lies in an inherent belief that the possible lack of validity in the context of the present study does not originate in the IELTS test itself but in the inappropriate use of its results. Rea-Dickins, Kiely and Yu's (2011) study will be reviewed to provide some background for the discussion of this potential problem.

3) Using IELTS scores as a 'hard criterion' for success

In an interesting research study on the use of IELTS scores prior to program admission, Rea-Dickins, Kiely & Yu (2011) explore the way IELTS scores are used as a strong -or 'hard'-evidence upon which decisions on the admission of international students to universities in the UK are made, and the subsequent impact on those students' post-admission academic profile. They put into question the appropriateness of the practices of decision-makers in using IELTS scores for program admissions that can affect the correlation between the those test scores and candidates' subsequent performance. Rea-Dickins, Kiely & Yu (2011) seem to question the validity of research studies claiming that IELTS has a low predictive validity, because those studies build their findings on factors that are sometimes extraneous to the test itself, like the erroneous use of test scores. The IELTS test scores in the case of the present study are used as

a 'hard criterion' for the graduation of the B.Ed. students with no clear justification of the relation between performance on the test and in the classroom.

In fact, Rea-Dickins, Kiely & Yu's research study focuses on factors which are beyond the test itself including the way stakeholders (decision-makers) use test scores, and the impact of wrong practices on international students' success and on their academic profile. They report that they gathered evidence that people involved in making decisions of this kind perceived IELTS test score as a true indication of a candidate's language ability, and used it as a 'hard' evidence. However, they clarify that the test providers were clear enough, in their message to the test users, about the necessity of considering scores of the four components of the test along with the overall score to see if they match a chosen program's linguistic demands, as well as taking factors other than language ability into consideration. They claim that many of the people involved in such decision-making have little knowledge about the test itself, about language proficiency, and about the various challenges that students face when they study a program in a second language at a tertiary institution. Their research findings reveal that, in contrast with the prescribed procedures of decision making which place language test scores at a lower rank than previous attainment and potential (Grade Point Average), experience at work (CV and references), commitment and learning purpose (reference letters), practice showed that those three factors were treated as 'soft' or flexible criteria because they were open to different interpretations. However, the IELTS scores in the case of international students were taken as a 'hard' evidence due to their gatekeeping minimum score. Moreover, Rea-Dickins, Kiely & Yu (2011, p.271) find that the non-language specialists who process international students' applications use IELTS scores in a 'rigid' way that does not involve much thinking, because they do not know much about the test. They reported that in some cases, those 'specialists' were not sure of the required IELTS band, and some others confused between IELTS and TOEFL scores. Only few of them who had some connection with people who had some knowledge and expertise in IELTS showed better understanding. When asked about their knowledge of the test and the interpretation of overall and sub-scores, many of them did not see a good reason to know more than the minimum score that tells them whether a candidate should be granted admission for a specific program or not. The researchers conclude 'there seems, thus, to be a pragmatic and also minimalist approach towards admissions decision-making: get it done, with a safe and 'hard' criterion' (2011, p.274).

2.6.4 Messick's unitary framework of validity

Throughout the history of language testing, validity has been associated with the general definition that a test should measure what it purports to measure (Cronbach & Meehl, 1955; Lado, 1961; Cronbach, 1971; Davies, 1990; Bachman & Palmer, 1996; and Weir, 2005). However, Messick's (1989) thorough examination and definition of the concept remains the main frame of reference for most researchers in the field of testing.

Messick defines validity as 'an integrated judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment' (1989, p.13). He identifies the interpretability, relevance, the utility of scores, the implication of scores for subsequent action, and their value for social consequences as the key issues of validity (*ibid*). He clarifies that validity is not concerned with the test itself but by the inferences derived from the test scores and the actions that follow. This is very important for the main question of the present study, because it articulates the principal issue related to the use of IELTS test scores as a graduation requirement for the bachelor of Education students in UAE. The choice of IELTS seems to be based on a shared assumption among several educational institutions in UAE that it is an internationally valid test that gained its validity from the fact that it is widely used around the world. However, the policy makers who made the decision to use it as an exit gate-keeper for the Bachelor of Education program at the Higher Colleges of Technology do not seem to have thought of validating the inferences and decisions based on the test scores. Therefore, inferences are based on IELTS classification of the candidates who achieve band 6 as competent users of English. Those inferences, have a direct impact on deciding who can be qualified as a teacher of English, which disregards the fact that a good speaker on the IELTS speaking test is not by definition a competent teacher unless their ability to accommodate their language to the level of their young learners in school is verified. Messick, also, states that validity is a matter of degree and not all or nothing. He distinguishes between validity as a property that is unified, though with different facets, and validation as a process of scientific inquiry that occurs in a political context.

He contends that 'validation embraces all of the experimental, statistical, and philosophical means by which hypotheses and scientific theories are evaluated.' (1989, p.14). Validity does

not account for discrete events or behaviours but encodes consistencies in behaviours or responses in the form of scores. However, scores are not just behavioural consistencies or attributes of persons but also judgmental consistencies and attributes of groups, situations, and objects (1989, 14). Messick emphasizes scores and test responses as sources of interpretation of behaviour consistencies rather than tests or instruments. He justifies his stance by indicating that test responses are ‘a function not only of the items, tasks, or stimulus conditions but of the persons responding and the context of measurement’ (1989, 14). Messick considers that the social psychology of the test context, the examinee’s environmental background and experiential history should be taken into account, which raises the question whether the interpretation of a test score should be context-dependent or generalized across contexts. As a result, Messick recommends that the role of context in interpreting test results and use should be investigated. He advocates that the context can affect the generalizability of the assessment interpretation more than its validity. ‘Thus one might seek and obtain generality of score meaning across different contexts, even though the attendant implications for action may be context-dependent by virtue of interactions operative in the particular group or setting’ (1989, p.15). This, in fact, is one of the objectives of the current study which seeks to investigate the generalizability of IELTS speaking scores in the specific context of classroom interaction where features of classroom discourse and teacher talk play a major role in shaping the subjects’ speaking performance and specifically in defining their lexical diversity. The question that motivates this research study is whether the subjects’ lexical diversity scores on the IELTS speaking component can be generalized to the classroom context to say that they have predictive validity.

Messick indicates that there is an array of basic sources of validity evidence, but he can identify only six, namely content, response, internal, external, differences in processes and structures, and the social consequences. He refuses to call them ‘types’ -as they are commonly called- but he explains that the relation between the evidence and the interpretation should determine the validation focus. He also clarifies that the varieties of evidence should not be regarded as alternatives but that they complement each other. That is why validity is a unitary concept that has different facets.

By reference to Anastasi, Cornbach and *APA Standards*, Messick (1989) indicates that since the 1950s, validity has been divided into three types which are content validity, predictive and concurrent criterion-related validity, and construct validity. He contends that together they cover all types of validity but ‘only social consequences of test interpretation and use are neglected’. He asserts that the evaluation of test validity and ‘whether it serves its intended function or purpose –requires evaluation of the intended or unintended social consequences of test interpretation and use’ (1989, p.84). He presents a unified framework of validity that has four facets: (1) Construct validity, (2) construct validity and relevance or utility, (3) value implications, and (4) social consequences.

In his framework, Messick (1989, p. 20) distinguishes between the evidential basis of test interpretation which is its construct validity, and the evidential basis of test use which is its construct validity that is supported by evidence of relevance to the purpose of the test and the setting where it is applied. In the case of the present study, the latter (the evidential basis of IELTS test use) is examined in order to see if there is enough evidence of relevance to the purpose of certifying graduating students as qualified teachers of English, and if it is relevant for the classroom context. On the other hand, he distinguishes between the consequential basis of test interpretation which is the evaluation of the ‘value implications of the construct label, of the theory underlying test interpretation, and of the ideologies in which the theory is embedded’ (1989, p.20), and the consequential basis of test use which is the evaluation of social potential and actual consequences of the used test. This consequential basis of IELTS test is going to be examined in the present study. The value implications of IELTS test score interpretation usually impact the judgment of a student teacher’s potential to be a good teacher or not. Also, the consequential basis of test use impacts on student teachers’ chances to graduate, or to suspend their year-four courses in order to get more IELTS practice until they achieve the required band.

While bearing in mind Messick’s focus on the relation between the evidence and the interpretation in determining the validation of a test, it is important to consider Kane’s argument-based validity (1992) which can be seen as a potential validation process of IELTS in the context of the present study.

2.6.5 Kane's argument-based validity

In the absence of an official statement that validates the use of IELTS scores as predictors of future student teachers' performance in Emirati schools, and considering the contingency that the findings of this research study might reveal a low level of predictive validity, it is worth considering argument-based validation that predefines the interpretation of scores in order to justify subsequent decisions. Kane's argument-based validity (1992) is a framework for intended score interpretation that has been confirmed as a useful framework by specialists in test validation like Chapelle, Enright & Jamieson (2010) who used it to develop a validity argument for the TOEFL test, and Chapelle (2012).

Kane (1992, p.527) associates validity with the interpretation assigned to test scores which involves an argument leading from test scores to score-based statements or decisions. In his recent article that clarifies validity argument, Kane (2012, p.34) states that an argument-based approach requires that claims should be stated and their credibility should be evaluated. He clarifies that this can be done through 'a chain or network of inferences and supporting assumptions that would get us from the test scores to the proposed interpretations and uses of the scores' (*Ibid*). That network of inferences and supporting assumptions, Kane (2012) calls *interpretive argument*, and the evaluation of its coherence and plausibility, he calls *validity argument*. He postulates that a 'test-score interpretation always involves an interpretive argument' (Kane 1992, p.527), with the test score as the basis, and the decisions made out of the interpretation as conclusions. In the case of the current study, the use of IELTS band 6 score as a requirement for student teacher graduation rests on an assumption that the inference from required band score is that the candidate is proficient enough in English language to be granted access to the final year and to graduate as a teacher of English. However, the absence of an argument that interprets the required band and the subsequent decision based on that band keeps the use of the score open to different interpretations which raises questions on its validity, as in the present study. Kane (2001) attributes the argument-based validation to Cronbach's work (1980, 1988) who structures validation around the social dimensions and the context of validity arguments which help to appraise the intended interpretation and use of test results. Kane's definition of the argument-based approach to validation states that it 'adopts the interpretive argument as the framework for collecting and presenting validity evidence and seeks to provide

convincing evidence for its inferences and assumptions, especially its most questionable assumptions' (1992, p.527). However, a 'complete statement of the claims included in the interpretation and the goals of any proposed test uses' (Kane 2001, p.329) is required for the evaluation of a suggested interpretation of test scores.

Inferences depend on different assumptions, which may or may not be credible. For example, inferences that are made out of test scores and are projected to non-test behaviour assume that there is a certain relationship between test behaviour and non-test behaviour. However, that 'assumed' relationship in the case of the current study is unclear which raised interest in investigating it. Also, inferences driven from test scores on theoretical constructs depend on some assumptions about the construct. According to Kane (1992) the interpretative argument cannot be verified in any 'absolute sense' because it is impossible to prove all the assumptions in the interpretive argument. Instead, he suggests demonstrating that the interpretive argument is 'highly plausible'. The interpretive argument is used as a framework for the collection and presentation of validity evidence and for providing convincing evidence to support inferences and assumptions, especially for controversial assumptions. Kane (2012, p.35) indicates that the interpretative argument makes the rationale for any score interpretation explicit so it 'can be evaluated in the validity argument'.

Kane (2012, p.37) postulates that the development of the interpretive argument goes hand in hand with the development of the assessment. He suggests to work backward from the conclusions and decisions, through the inferences, to the assessment. Thinking of the context of the present study while echoing Kane's example (*Ibid*), the procedure should start with identifying the target performance in the context of the classroom, then designing assessment tasks for that context. After that, criteria for the evaluation of performance is developed. The interpretive argument starts from identifying the interpretation of interest which corresponds to the level of language proficiency that is appropriate for the classroom. Then, while moving backward from decisions to grant student teachers access to the final year of their program or not, to observation of their performance on the assessment, a 'chain' of inferences related to their English language proficiency for a classroom context is developed based on well identified assumptions.

Kane (1992, p.528) identifies three general criteria for the evaluation of practical arguments:

1- Clarity of the argument: Inferences and assumptions should be stated explicitly and in ‘enough detail’.

2- Coherence of the argument: It should be consistent with the ‘rules of logic and mathematics’, and/ or theory.

3- Plausibility of assumptions: They should have enough evidence that supports them.

For Kane (1992), some interpretive arguments can be problematic if they miss any of the criteria, but they can still be monitored and corrected. However, ‘hidden assumptions’ can be serious problems because no evidence is gathered to support them, which is the case in the context of the current study that caused this research to take place.

2.7 Conclusion

Drawing on this literature review, the thesis takes as a point of departure an understanding of language proficiency through three models that have clear connections with the main research argument stating that IELTS scores may not be good predictors of student teachers’ performance in a classroom situation. Bachman and Palmer’s (1996) framework is taken to examine the correspondence between IELTS speaking test as a test task and classroom teaching as a language use task. The evaluation of that correspondence will be framed within the relevance of the criterion (IELTS score) and its predictive validity. Cummins’s (2000) framework will provide another angle for the examination of that correspondence which focuses on the cognitive and contextual factors involved in each situation. The evaluation of that correspondence will be interpreted in terms of the cognitive and the context validity of the test scores. Canale and Swain’s (1981) framework of communicative competence represents the third edge of this triangulated review of language proficiency. It will help to examine the correspondence between the subjects’ performance on IELTS and their performance in a classroom situation, especially with reference to their sociolinguistic and strategic competences. This will be achieved by referring to Bell’s audience design (1984) which provides a useful framework to analyse the addressee’s impact on the speaker’s performance. This thesis sets out to begin to address this issue by examining teacher talk in order to identify the unique architecture of classroom interaction (Seedhouse, 2004) and its relevant features that impact lexical diversity, and therefore language proficiency in the classroom.

The main concern in this thesis is the use of individuals' language proficiency scores to predict their ability to perform successfully in a language teaching classroom in the UAE. With respect to testing, previous research has revealed that high levels of language proficiency correlate with high levels of lexical proficiency, and high scores on lexical diversity are good indicators of lexical proficiency. However, what is less clear is whether a good score in lexical diversity on the IELTS speaking test can predict the ability to use it appropriately according to context and addressees, and how one might set about evaluating an individual's ability to modulate their lexical diversity in response to classroom needs.

Chapter 3: Methodology

3.1 Introduction

In this chapter, I will give a detailed description of the methodology followed in this research study. The focus of the study is developed in section (3.2), and the research questions are presented in section (3.3). Section (3.4) presents the research paradigm and epistemology, followed by the methodological approach in section (3.5) and the research design in section (3.6). In section (3.7) a detailed description of the data collection instruments is presented, and in section (3.8) data collection procedures are explained. Data analysis procedures and tools are described in section (3.9). In sections (3.10) and (3.11), research validity and reliability, and research ethics are developed respectively. Finally, methodological issues were discussed in section (3.12), and a reflection on the research process is presented in section (3.13) to conclude this chapter.

3.2 Focus of the study

The current study focuses on investigating the predictive validity of IELTS scores for a teacher education program in the UAE. It questions whether using IELTS scores as a gatekeeping device blocking those who fail to get the required band from the final year of studies and graduation is an appropriate measure. The study focuses on the speaking skill of the subjects as it is the most performed skill in the classroom. It compares the subjects' speaking performances on the test with their performances in the classroom to gauge the strength of the relationship between the two tasks and to find whether the test scores can be used as valid predictors of the classroom performance or not. A consequential explanatory mixed methods design is used whereby quantitative and qualitative methods in data collection, analysis and interpretation are combined. Quantitative data gauges the significance of the relationship between the test scores and students' performance in the classroom, and qualitative data follows up on the quantitative findings to provide insight into the factors that explain the strength or the weakness of that relationship. The findings will inform the discussion whether the use of IELTS scores is valid

to make high stake decisions related to the graduation of the Bachelor of Education students in UAE.

3.3 Research questions

The main question of this research study is:

Can IELTS test scores make accurate predictions of student teachers' performance in the L2 classroom?

The sub-questions are:

- 1) Do student teachers display similar lexical diversity on the IELTS speaking test and in the classroom?
- 2) Do IELTS scores match college mentors' grades for student teachers on teaching practicum?
- 3) How does classroom interaction affect student teachers' lexical diversity scores?

I shall explore the relevance of each of these sub-questions to answering the main research question in the rest of this chapter

3.4 Research paradigm and epistemology

Following the validity issue raised by this research study and its significant impact on the B.Ed. students' graduation opportunities, mixed methods research design combining quantitative and qualitative methods is chosen based on philosophical and paradigmatic considerations that will be explained in this section. The quantitative research method is used first to provide generalizable findings that can apply to the six campuses where the program is run. Due to its strength as an indicator of language proficiency (section 2.3.1), lexical diversity is used to compare between the subjects' performance on IELTS test and their performance in the classroom. Statistical data obtained from the comparison is analysed and interpreted to answer the first sub-question on whether student teachers display the same lexical diversity in both

situations or not. A second statistical comparison is conducted to gauge the strength of correlation between IELTS scores and Mentoring College Teachers' (MCT) scores of the subjects' performance on teaching practicum. The quantitative data is analysed and interpreted to answer the second sub-question.

However, this quantitative method of research which is based on statistical findings and analysis can only confirm or disconfirm the original hypothesis of this research that there is no correspondence between IELTS scores and students' language proficiency in a teaching context. In other words, this method does not provide a detailed explanation for the findings to conclude that the existing form of testing is valid and justifiable, therefore worth maintaining, or has a weak validity value that it needs to be reconsidered and possibly substituted by another assessment tool. A qualitative research method is thus needed to examine the scripts of the participants' speaking performances in order to capture details that characterise each context and how they affect the subjects' lexical diversity. Ultimately, those details provide qualitative data that can help to answer the third sub-question on the characteristics of classroom interaction and their impact on student teachers' language proficiency.

Despite the 'paradigm debate' (Creswell and Clark 2011, p.25) that discusses whether quantitative and qualitative paradigms can be reconciled because they are linked with different philosophical assumptions, the growing use of mixed methods research over the last decade has demonstrated the need for such a method to address 'the complexity of our research problems' (Creswell and Clark 2011, p.21). The philosophical assumption that underpins the combination of those two different paradigms is known as 'pragmatism'. Tashakkori and Teddlie (2003, p.713) define pragmatism as a 'deconstructive paradigm' that focuses on 'what works' to help researchers answer their research questions instead of restricting them to either the quantitative or the qualitative paradigm, because according to 'traditional dualism' they cannot be reconciled. In this research study, in order to achieve a clear understanding of the relationship between IELTS test scores and their interpretation for the qualification of B.Ed. students as teachers of English in UAE schools, a pragmatic approach is adopted where quantitative tools are used to confirm or disconfirm the existence of a relationship, and qualitative tools are used to explore the reasons behind the quantitative findings and to develop a better understanding of the nature of that relationship. A mixed methods approach is opted for based on a belief that the

nature of knowledge and the way it can be explained and justified should not be confined to one model of thought and its related approach to research. Mixing paradigms and mixing methods in a careful way that draws on a combination of their strengths can help understand phenomena under investigation in a better way and can make a researcher more confident of their findings and interpretations.

3.5 Methodological approach

Teddle and Tashakkori (2009, p.21) make a clear distinction between methodology and research methods. They define research methodology as ‘a broad approach to scientific inquiry’ whereas research methods as ‘specific strategies and procedures for implementing research’. In this research study, the methodological approach is guided by the following principles:

- 1- The research questions are asked and answered in a way that helps to gauge the strength of the relationship between IELTS scores and decisions based on them that qualify or disqualify B.Ed. students for graduation in the context of the present study. The questions should also provide a detailed explanation that justifies the level of strength of that relationship in order to make conclusions about the validity of using IELTS scores as a graduation requirement.
- 2- To answer the guiding questions in a way that maximizes understanding of the relationship between the two main variables of the research study, a pragmatic stance is taken where quantitative and qualitative worldviews could be combined at any stage of the study. This approach stems from a widely-shared view that we should focus on what works, rather than dogmatically pursue one particular method.
- 3- A preference for a *sequential mixed design* where the quantitative phase of the research study precedes the qualitative phase in order to provide a quantitative evaluation of the strength and direction of the relationship between the two main variables, then a qualitative explanation of the reasons why it is at that level of strength and the significance for the main question of the study. Qualitative analysis of the relationship between the two main variables in this research study ‘expand(s) on the initial

understanding gained from the quantitative analysis' (Teddlie & Tashakkori 2009, p.170).

- 4- Sampling depends on the availability of volunteers in a context where sociocultural factors play an important role in limiting the size of the population and its representativeness. *Convenience sampling* (Teddlie & Tashakkori 2009, p.170), i.e. utilizing participants who are immediately available, helps to engage the maximum of possible volunteers who manage to get their parents'/guardians consent to take part in the study.
- 5- Data collection and data analysis strategies follow an integrative eclectic approach with disregard to any claimed boundaries between the quantitative and qualitative paradigms in order to answer the research questions.
- 6- A solid process of making inferences produces quality and transferability/generalizability of the research conclusions. Inferences reflect the efficiency of the research design in attaining the research aims.

3.6 Research design

Based on the main problem identified in the focus of the study (section 3.2), and on the research paradigm and the methodological approach explained in sections (3.4) and (3.5) respectively, a *mixed methods* research design is chosen to answer the main question. Creswell (2015, p.2) considers mixed methods a method that focuses on data collection, analysis and interpretation. He defines mixed methods in the following way:

An approach to research in the social, behavioral (*sic*), and health sciences in which the investigator gathers both quantitative (closed-ended) and qualitative (open-ended) data, integrates the two, and then draws interpretations based on the combined strengths of both sets of data to understand research problems. (*Ibid*)

The first two sub-questions of this research study (section 3.4) justify the use of a quantitative method in order to gauge the level of the relationship (if any) between the lexical diversity displayed by the subjects on the IELTS speaking test and of their teaching in the classroom. The

quantitative method is also required to see if there is any statistical correlation between the subjects' scores on the IELTS speaking test and their college mentors' scores of their teaching practicum. The quantitative strand is the first phase of the research study to be able to qualify the relationship of IELTS test scores and the subjects' language performance in teaching as strong, weak, or non-existent. The subsequent qualitative phase is the second phase which helps to understand the factors that make the strength of the relationship significant or insignificant. It is an important phase because it adds clarifying details to the answers provided by the quantitative phase in order to answer the main question which investigates the ability of IELTS scores to make accurate predictions of student teachers' performance in the classroom in order to use them for a high stake decisions related to the subjects' eligibility to graduate as teachers of English in UAE schools. An *explanatory sequential design* (Creswell & Clark 2011, p.81) is used with a qualitative phase that follows up on the quantitative phase findings to provide in-depth explanation of the characteristics of classroom interaction and how it affects student teachers' language proficiency in the classroom. The quantitative phase in this sense is an initial stage of the research study to find out whether the relationship between the two variables is significant or not, whereas the qualitative phase is an explanatory stage that helps to understand why the relationship is significant or not and therefore to have clear arguments to confirm or disconfirm the validity of using IELTS scores as a requirement of graduation for the Bachelor of Education students in UAE.

3.7 Data Collection instruments

The data collection instruments presented in sections (3.7.1) and (3.7.2) are used to collect quantitative and qualitative data that help to implement the explanatory sequential research design explained above. At a first phase a mock IELTS speaking test is conducted for the participating subjects and recordings of their teaching in the classroom are collected to run a statistical comparison between their lexical diversity scores in both contexts in order to find out if they correlate or not. The mock speaking test scores that are assigned by the examiners are used for a comparison with the college mentors' scores of the subjects' performance on teaching practicum. The aim of this second comparison is to see if by any means the overall speaking score on the test would agree with the scores on teaching practicum. The significance (or

insignificance) of the relationship revealed by the statistical phase is then explained by conducting a qualitative analysis of the subjects' interaction in the classroom using the collected recordings to identify the factors that play an important role in making their speaking in the classroom similar to or different from speaking on the test.

3.7.1 The 'mock' IELTS speaking test

The researcher selected four different IELTS speaking tests from different published resources that were used by the examiners for the mock test:

1- Test No 1:

Part 1: about the place where the candidate grew up and how it has changed.

Part 2: about a sports event that the candidate enjoyed watching.

Part 3: about the cost of watching sports events, the difference between watching sports events in live and on TV, and sports celebrities.

Resource: *Official IELTS Practice Materials - 2007* (appendix 2).

2- Test No 2

Part 1: about the place where the candidate lives, sports and cooking.

Part 2: about a shop that the candidate enjoys going to

Part 3: about shopping and consumerism

Resource: *IELTS Express Intermediate - 2006* (appendix 3).

3- Test No 3

Part 1: about where the candidate lives, friends and eating habits

Part 2: about a film that the candidate found interesting.

Part 3: about social events

Resource: *IELTS Practice Tests - 2006* (appendix 4).

4- Test No 4

Part 1: about studying, weather and weddings

Part 2: about a lucky person that the candidate knows

Part 3: about superstition

Resource: *Step up to IELTS* - 2004 (appendix 5).

3.7.2 The class recordings

The 27 subjects who took part in the interviews were then asked to record one of their classes while they were on teaching practicum placement (TP) in schools. Student teachers were given the instruction to record any class they were to give during their TP without any specific choice of the theme or the language focus of the class. The submitted recordings varied in length according to the grade level that the student teachers were teaching. Primary classes (grades 1-5) usually last between 30 and 40 minutes, whereas middle school classes (grades 6-9) last between 45 and 55 minutes.

3.8 Data collection procedures

3.8.1 The setting

The data for the main study is collected from Emirati student teachers studying in the Bachelor of Education Program (B.Ed. Program) at Fujairah Women's College (FWC), The Higher Colleges of Technology (HCT), United Arab Emirates (UAE). The HCT is the largest higher education institution in the UAE that enrolls more than 18000 students, and it is based on 17 men's and women's campuses.

The Bachelor of Education Program, English Language Teaching in School (B.Ed-ELTS) is a full-time four-year program that was established in collaboration with the School of Education of the University of Melbourne, Australia. It has received accreditation from the same university since it was first created in 2000, and it continues to do so. The program aims to produce qualified Emirati teachers of English who are eligible for jobs offered by the UAE Ministry of Education.

Students join the program after studying for one year in the foundation program and after a selection process based on their scores on the IELTS test. The minimum entry requirement is an overall band 5 with no skill band less than 5. However, some students are granted a direct entry from high school after taking a 'challenge test' and attaining IELTS band 5.

Along with the education courses, the program offers an English language studies course every semester that focuses on developing the students' English language skills to help them achieve IELTS band 6 by the end of the fourth semester. Indeed, by the end of their year 2, students are asked to take the IELTS test and provide an official IELTS report showing that they achieved the required band. Students who fail to do so are given a chance to pass to year 3 if they passed all the year 2 coursework and final assessments. In year 3, those students keep retaking the IELTS test until they do achieve an overall band 6 and no skill band less than 5.5. However, if they fail to do so by the end of year 3 they are barred from enrolling on to year 4 even if they passed all the year 3 coursework assessments. It is a fundamental condition that no student can be admitted to the final year of the B.Ed. degree if they do not have band 6 in IELTS. However, such a scenario rarely happens.

All the students in the B.Ed. program go on a four-week Teaching Practicum placement (TP) every semester in which the focus is on developing specific teaching skills. They are assessed through interviews, portfolios, the Mentoring College Teacher's (MCT) report, the Mentoring School Teacher's (henceforth MST) report, and their reflections on assigned tasks. The year 3 TP assessments are as shown in table (3.1) below.

MCT grades are assigned and submitted with a report that they write after attending two observation lessons with each student, and after discussing students' performance on teaching practicum (TP) with the students themselves, the mentoring school-teacher and the principal of the school. The report is divided into five main sections, each one covering one of the following competencies:

- 1- Professionalism and understanding
- 2- Planning for learning
- 3- Implementing and managing learning
- 4- Monitoring and assessment
- 5- Reflection

The descriptors for each competency are provided in appendix (1).

After writing their reports, MCTs assign an overall TP score out of 100. This practice (writing a report and assigning a score) is a common practice across the six campuses where the B.Ed. program is run.

Table 3.1 *B.Ed. year-three Teaching Practicum assessment*

First Semester Assessments			Second Semester Assessments		
1	MCT's Report	45%	1	MCT's Report	45%
2	MST's Report	15%	2	MST's Report	15%
3	Group project (planning a whole unit of instruction)	40%	3	Peer assessment	20%
			4	An online journal	20%

3.8.2 Sampling and permissions

The process started by explaining the research study and the aim behind it to year 3 and 4 students to know their level of interest. The choice of year 3 and 4 students is based on the fact that they are the ones who have taken IELTS test by the end of year 2 as mentioned in the previous section. They represent the sample of the B.Ed. students to whom the policy of getting IELTS band 6 to graduate is applied. No student was excluded from the intended sample including those who have failed to achieve the required band and who were still retaking the test. The purpose is to use all the available sample in order to avoid the 'truncated sample' issue associated with research studies on the predictive validity, as discussed in section (2.6.3).

As soon as the college Research Board gave approval to start the research, parents/ guardians' consent forms were given to all the students who showed interest in participating (appendix 136). Though most of the participants were above the age of twenty, in the context of the present study women must get their parents or their guardians to sign for them following sociocultural norms. Upon receiving the consent forms back from the initial candidates who expressed interest, the researcher found that the sample size was clearly reduced to 27 out of 42 due to some parents' and guardians' refusal to grant their dependents permission to participate in the study.

Soon after the collection of consent forms, the researcher asked five certified Cambridge ESOL examiners who worked as faculty in the foundation program at the same college to conduct a mock IELTS speaking test under similar IELTS conditions and procedures.

The next stage after conducting the IELTS ‘mock’ test was asking the same students to record one of their classes when they were on teaching practicum placement. The researcher provided recording devices and CDs but most of the subjects preferred to use the audio recording software available on their laptops.

All the recordings, both of the mock IELTS speaking test and the classes were then saved on CDs and flash memory discs and locked in a drawer at the researcher’s office.

3.9 Data analysis procedures

To conduct data analysis, the researcher started by transcribing the recorded speaking test and the class teaching sessions. An identification code was assigned for each participant instead of their names for the sake of anonymity. For the quantitative phase of the research, the transcribed texts were converted into CHAT format (section 3.9.1-1 below) to facilitate the computation of the participants’ lexical diversity using the tool VOCD (3.9.1-2) in CLAN program (3.9.1-1). In order to answer the first research sub-question on whether the participants display the same lexical diversity on the test as in the classroom, a correlation between the two variables was run using the Statistical Program for the Social Sciences (SPSS). Then, the same tool (SPSS) was used to help answer the second sub-question on whether there is a correlation between the students’ scores on the IELTS speaking test and their college mentors’ scores by running a correlation between the two sets of scores.

To provide more statistical insight into the findings of the correlations, a second quantitative test was used. *WordSmith tools* (section 3.9.1-3) were used to provide more statistics and detailed information about the use of content words in each context. First, *Word Frequency* tool was used to generate word frequency lists of each context to examine similarities and differences of high frequency content words produced in a test situation and in the classroom with the aim of telling whether one context can predict the other. Second, *WordSmith* was used to run

concordances to examine the sentential context of identified high frequency words in order to gain understanding of the impact of context on the choice of lexical items. A third tool of WordSmith that was used is *key word* tool to see how the context of the speaking event plays a role in determining the subjects' most common words that are identified in relation with their overall speech.

Using corpus linguistics (3.9.2-2) which includes all the transcribed data from the recorded mock IELTS speaking test and the classroom teaching sessions, WordSmith tools provided data that is useful for both quantitative and qualitative purposes. Tables of word frequency lists of both the test situation and the classroom and graphs of percentages of parts of speech (nouns, verbs, adjectives, and adverbs) helped to see in more detail similarities and differences of words generated in each situation. Examining word lemmas and their collocates in the classroom using concordances prepared for the qualitative data analysis where reference was made to specific examples of the subjects' use of words and the context of their use. In fact, some examples of concordances and key words were useful to answer the third question of the research which explores the impact of classroom interaction on the subjects' lexical diversity.

Based on the findings of the quantitative phase, a qualitative phase that is focused on selected samples of the subjects' speech in the classroom was conducted. The analysis followed Seedhouse's framework of the architecture of classroom interaction (3.9.2-1) to identify features of classroom interaction that characterise the subjects' talk in the classroom and that control the diversity of their lexical items. Then, both quantitative data analysis drawn from corpus linguistics, and qualitative data analysis based on Seedhouse's framework are combined following Walsh's approach (CLCA) to the analysis of classroom interaction (3.9.2-3). The purpose was to provide detailed analysis of the effect of classroom interaction on the lexical diversity of the subjects.

3.9.1 Quantitative analysis tools

1) CHAT and CLAN

CHAT and CLAN are two computational tools of the Child Language Data Exchange System (CHILDES) that are used in this study to transcribe the recorded interviews and the classes, and to compute the lexical diversity of the subjects' speeches in both situations. CHILDES is a

widely-used corpus offering ‘a set of computational tools designed to increase the reliability of transcriptions, automate the process of data analysis, and facilitate the sharing of transcript data’ (MacWinney 2010, p.5).

a- CHAT

Codes for Human Analysis of Transcripts (CHAT) is a transcription tool that ‘provides a standardised format for producing computerised transcripts of face-to-face conversational interactions’ (McWinney 2010, p.14). It offers a model of transcription that follows specific conventions in transcribing and coding speech in a way that warrants consistency especially when the number and length of transcribed talks are relatively high as in the current research study. Moreover, the coding system provided by CHAT facilitates the task of researchers by providing options that can help them to focus on specific parts of the transcribed speech that are meant to be analysed at a subsequent stage by CLAN. One of the very useful codes that were used in transcribing the subjects’ interviews and classes in this study is the ampersand (&) which excludes any word that comes after it from the analysis. This code was very effective in eliminating all the words that were judged unnecessary for the analysis of the lexical diversity of the subjects’ speech. These included:

- 1- proper names of people like the names of the examiners, the candidates, and the schools of the student teachers.
- 2- names of countries, cities and towns like UAE, France, Fujairah, Dubai, Khorfakan, etc.
- 3- acronyms like *IETLS*, FWC, KG, etc.
- 4- the exclamation ‘OK’ due to the fact that it was excessively used by the subjects which could skew the computation of lexical diversity.

It is noteworthy that all the contracted forms like ‘isn’t’, ‘didn’t’, and ‘let’s’ were transcribed in their full forms (is not, did not, and let us) after noticing that the short forms were not always recognised by CLAN in the analysis stage, and were not classified according to their morphosyntactic categories but were assigned a question mark (?) that excluded them from the computation.

b- CLAN

Computerised Language ANalysis (CLAN) is a program offered by CHILDES that allows the automatic analysis of large data that is transcribed in CHAT format. This program gives the user a number of analytic commands that can serve different computational purposes of any analysis. Some of those commands are the frequency command (FREQ) which is used to compute the frequency of words in a file and the VOC-D command that is used in this research study to compute the vocabulary diversity of the subjects' speeches both on the speaking test and in class situation (MacWinney 2010).

CLAN uses a general library (CLAN lib) that functions as a reference directory. Within the CLAN lib there is the MOR lib that functions as a reference directory for the analysis of grammar and lexicon. As the MOR command is run on the CHAT file, a MOR tier is generated below the main tier that shows the morpho-syntactic analysis. The command breaks each word into morphemes and assigns a syntactic category to them. The MOR program disregards the context of the processed language and assigns all the identified words all the syntactic categories and morphological types that are possible as shown in the example below:

```
*AMN:      I am from &uae .
```

```
%mor: pro:sub|I v:cop|be&1S^aux|be&1S prep|from .
```

Based on this example, the analysis of the second word (am) is ambiguous as MOR program suggests two possible categories separated by the caret (^). Consequently, it is not clear whether 'be' is considered as a verb or an auxiliary. For this reason, CLAN provides another command, which is the POST command, to disambiguate similar analyses. As we run POST command the result shows as follows:

```
*AMN:      I am from &uae .
```

```
%mor: pro:sub|I v:cop|be&1S prep|from .
```

So, the ambiguity is cleared as (am) is categorised as a copular verb.

In the case of the current study, the computation of VOC-D for each subject went through the following stages:

- 1- Transcribing all the recordings in CHAT text format.
- 2- Running a check over the compatibility of the transcribed texts with CHAT rules to ensure that all the words are recognised by the program.
- 3- Checking that all the words exist in the MOR lib directory by running the command:
mor +xl @
- 4- Running the MOR command onto the CHAT files of each subject for a morpho-syntactic analysis.
- 5- Running the POST command to disambiguate the unclear categorisation of some morphemes after running the MOR command.
- 6- Running the VOC-D command to compute the lexical diversity of the subjects in each text.

CLAN offers an array of options based on automatic morpho-syntactic coding to help the researchers get precise results from the computation performed by the program. It presents special codes to choose the appropriate command in a CHAT file, to select specific lines (tiers), and to include or exclude some features of the analysis.

2) VOC-D

VOC-D is a vocabulary diversity index that was created as an alternative to Type-Token Ratio (TTR) which has always been criticised for lack of reliability because it shows clear sensitivity to text length, and therefore produces misleading findings. The main flaw of the TTR measure is that it relies on a linear calculation of the ratio of word types (Type) by the number of their occurrences (Tokens) in a text. Usually, the obtained result is very much affected by the text length as the value of TTR declines with the increasing number of tokens and the systematic decrease in the number of types. As the text gets longer and longer, the chances that the language user produces new types diminish to the extent that some language users exhaust all their vocabulary types and keep repeating the same types. When this happens, the tokens size increases considerably at the expense of the type size and therefore skews the results and distorts the final findings. (Malvern & Richards, 1997; McKee, Malvern & Richards, 2000; Vermeer 2000; Malvern & Richards 2002; Daller, Van Hout & Treffers-Daller 2003; Duran et al. 2004; McCarthy & Jarvis 2007 & 2010).

Unlike TTR, VOC-D does not rely on linear counting but on random sampling that uses 100 random samples without replacement of 35 tokens from the script under investigation and calculates the mean of their TTR values, then repeats the same procedure for samples from 36 to 50. Based on the mean of each of the 100 samples, an empirical TTR curve is created. Then, a probability formula using \mathcal{D} coefficient is used to create an ideal or theoretical curve that best fits the empirical TTR curve by the least square difference. The best-fitting \mathcal{D} is then taken as the D value. To depress the effect of randomised choice of tokens and to increase consistency, the procedure is run three times to come up with a final D value that represents the best fitting between the empirical and the theoretical curves (McCarthy and Jarvis 2010, p.383).

MacKee, Malvern & Richards (2000) give three clear reasons why this index is superior to other indices:

- 1- It does not depend on the length of the text as it is the case with TTR. The random sampling and the best fit between the theoretical and the empirical curves solve the problem of the token size that increases with the text length.
- 2- Unlike some models which tried to solve the problem of text length by standardizing the number of utterances, this index uses all the available text.
- 3- It gives a better idea about the variety of TTRs across the text as different samples have different token sizes.

These strengths and the success of this index in overcoming the flagrant flaw of the TTR index justify why it was selected and used in this study for the computation of the subjects' lexical diversity for both the IELTS speaking test and the class teaching situation.

Despite their criticism of VOC-D, McCarthy & Jarvis (2007) confirmed its reliability and its success in becoming the first choice for both researchers and students who are interested in computing lexical diversity:

vocd appears to be steadily becoming the LD index of choice for researchers and students alike. And indeed, initial results of *vocd* appear promising (...) with some researchers (...) appearing already to be treating *vocd* as the industry standard (McCarthy & Jarvis 2007, p.461).

2) *WordSmith Tools*

WordSmith Tools is ‘an integrated suite of programs for looking at how words behave in texts’ (Scott 2010, p.2). It is mainly used in corpus linguistics to see how individual words are used in a text. The software package has three main tools:

- 1) *Concord tool*: it is a program that makes concordances using a plain text to give the context in which a selected word is used. It gives information about collocates of the selected word as in table (3.2)

3.2 *A concordance sample*

N	Concordance	Word #	Sent. #	Sent. Pos.	Para. #
1	ur first draft . I want you to check your mistakes and then	912	49	11.11%	151
2	ors . write your final draft . check your mistakes because w	929	49	48.89%	154
3	shed ? look at your mistakes . check your mistakes . wait wa	878	47	31.25%	144
4	nt you to use your red pen and check the spelling and grammar	859	46	40.00%	138
5	elling and grammar mistakes . check them . together togethe	864	46	60.00%	140

- 2) *Word list tool*: it is a program that generates word lists for a chosen text and orders them according to their frequency (Scott, 2010). It gives the different word-form counts as in table (3.3).

Table 3.3 *A word list sample*

N	Word	Freq.	%	Texts	Lemmas
1	BE	102	0.210304946	1	be[4] are[3] is[75] was[20]
2	YOU	79	2.733964205	1	you[52] your[26] yourself[1]
3	VERY	75	3.732912779	1	very[71] wadha[3] wadha's[1]
4	THE	72	3.785488844	1	
5	A	57	2.996845484	1	
6	OK	49	2.576235533	1	
7	AND	45	2.365930557	1	
8	LOOK	45	2.365930557	1	
9	THAN	44	2.313354254	1	
10	WHAT	40	2.103049517	1	

- 3) *Key-word tool*: It is a program that helps to identify key words in a text. ‘It compares the words in the text with a reference set of words usually taken from a large corpus of text. Any word which is found to be outstanding in its frequency in the text is considered ‘key’ (Scott 2010, p.5).

This is an example in table 3.4:

Table 3.4 *A key-word sample*

N	Key word	Freq.	%	Texts
1.	IS	23.00	2.84	1
2.	BOYS	20.00	2.47	1
3.	WANT	50.00	6.18	1
4.	MOHAMED	15.00	1.85	1
5.	AHMAD	12.00	1.48	1
6.	DRAW	15.00	1.85	1
7.	ABDULLA	6.00	0.74	1
8.	QUICKLY	6.00	0.74	1
9.	HAND	18.00	2.22	1
10.	BEEFBURGER	4.00	0.49	1

3.9.2 Qualitative analysis tools

The qualitative phase of the research which is an explanatory phase as the research design indicates in section (3.6) focuses on exploring features of classroom interaction and their impact on lexical diversity scores. The aim is to provide explanations for the quantitative findings and to answer the third sub-question of the study on the effect of classroom interaction on the lexical diversity of the subjects. Unlike the quantitative phase in which comparisons between IELTS scores and teaching scores were drawn, in this qualitative phase the analysis focuses only on the classroom data. This choice is made based on the fact that the main argument of this research study is that IELTS scores might not be to the right indicators of student teachers' English language proficiency in the classroom. This hypothesis is built on the researcher's assumption that classroom interaction has special features that cannot be found in any other kinds of interaction due to the uniqueness of the classroom context. Therefore, exploring those special features in the subjects' transcribed classroom data and analysing how they impact their lexical diversity provides an answer to the third sub-question and contributes qualitative answer to the main question of the study with no clear need to analyse IELTS transcribed data.

Three main tools are used for this qualitative analysis phase, (1) Seedhouse's (2004) model of analysis of the L2 classroom interaction, (2) corpus linguistics, and (3) Walsh's (2011) combined approach of corpus linguistics and conversation analysis.

1) *Seedhouse's framework for the analysis of the interactional architecture in the L2 classroom*

Seedhouse's (2004) framework for the analysis of classroom interaction in the L2 classroom is specifically derived from the second language classroom, which makes it very appropriate for the present study. Seedhouse (2004) adopts a conversation analysis approach to classroom interaction analysis that takes into consideration the specific features of the context. He conducts a study based on a data base that he collected from seven different sources covering a wide variety of L2 contexts and he demonstrates that conversation in the L2 classroom follows some rational order and embraces the principle of complementarity between randomness and rational design, simplicity and complexity, and homogeneity and heterogeneity.

According to Seedhouse (2004, p.12), Conversation Analysis (henceforth CA) 'studies the organisation and order of social action in interaction' and the role of a researcher is to analyse that organisation from an emic perspective that helps to disclose the mechanisms underlying them. An *emic* perspective studies behaviour from inside a system in a way that it considers the participant's viewpoint, whereas an *etic* perspective studies behaviour from outside a system that takes the analyst's viewpoint (Pike in Seedhouse 2004). Seedhouse identifies four principles of conversation analysis:

- 1- Interaction is rationally organised as there is order at all points, and talk is 'systematically organized, deeply ordered, and methodic' (Seedhouse 2004, p.14).
- 2- Contributions are shaped by the environment in which they are produced (*context-shaped*), and once they are produced they participate in shaping subsequent contributions (*context-renewing*).
- 3- There should be no instance where the order of detail 'can be dismissed a priori as disorderly, accidental, or irrelevant' (Seedhouse 2004, p.14).
- 4- The analysis of interaction follows a bottom-up approach and is guided by the available data. No theoretical assumptions, knowledge of background, or contextual details are considered in advance.

These four principles will, in fact, help to examine classroom transcripts for the subjects in the present study to try to understand the way they use their lexical diversity in accordance with the context of teaching young learners.

Seedhouse identifies four types of interactional organization:

- 1- *Adjacency pairs*: they are utterances that are paired up in a way that once the first part is produced, the second part becomes ‘conditionally relevant’ (Seedhouse 2004, p. 17), though not necessarily provided.
- 2- *Preference organization*: it is not what social actors prefer to do in interaction but rather what they should do to achieve their social goals through interaction. This happens as a result of the social norm of interaction which is affiliative. Interactants try to achieve reciprocity of perspectives. There is an intrinsic bias, in the organization of talk, towards maintaining a strong relation between interactants to avoid conflict.
- 3- *Turn-taking*: With reference to Sacks et al. (1974), Seedhouse indicates that turn-taking is governed by mechanisms whereby interactants have a set of norms that they can choose among. There are turn-constructive units (TCUs) that can be sentences, clauses, words, or nonverbal cues. A TCU is a social rather than a linguistic concept, at the end of which there is turn relevance place (TRP), which is a point of interaction at which a listener expects a speaker to end their turn. For Seedhouse, taking a TCU or a turn as a unit of analysis is a wrong practice because it ‘is a social action that is embedded in a sequential environment’ (2004, p.33). Instead, Seedhouse suggests that CA endorses a holistic approach to analysis that is not prescriptive as in linguistics.
- 4- *Repair*: Seedhouse (2004, p.34) defines repair as the treatment of trouble that impedes communication between participants. With reference to Schegloff, Jefferson & Sacks (1977), he identifies four ‘trajectories’ of repair:
 - (a) Self-initiated self-repair
 - (b) Self-initiated other-repair
 - (c) Other-initiated self-repair
 - (d) Other-initiated other-repair

On the other hand, Seedhouse (2004, p.37) indicates that the four types of interactional organization mentioned above should not be taken as ‘rules, units, or coding schemes’ as in a

linguistic model but as ‘normative resources’ that interactants use to share the social meaning of their actions.

In his discussion of the methodology for the analysis of L2 classroom interaction, Seedhouse indicates that his description of the interactional architecture outlined above provides a ‘ready-made emic analytical procedure’ (2004, p.195).

Like the teacher, the analyst compares the linguistic forms and patterns of interaction produced by the students with the pedagogical focus introduced at an earlier stage by the teacher and conducts an analysis which is based on the degree of match or mismatch. A key factor for the success of this procedure is the identification of the pedagogical focus.

Seedhouse indicates that the institutional context is talked into being by the participants when an institutional focus is introduced. Only an emic perspective can reveal whether participants talk the institution in or out of being at any moment of their interaction. The institutional context is defined by ‘an overarching variety of discourse which is suited to the overarching institutional aim’ (Seedhouse 2004, p.204). The L2 classroom context is, then, regarded as ‘the actualization of the reflexive relationship between pedagogical focus and interactional organization’ (Seedhouse 2004, p.205). That actualization would differ from one instance of interaction to another which makes the L2 classroom context dynamic and variant. However, Seedhouse indicates that the L2 classroom context is only one component of the classroom interactional architecture and just one aspect of the complex and reflexive relationship between pedagogy and interaction.

In CA, context is established from an emic perspective that identifies the elements of context which are relevant to the participants. Structural organizations, like turn-taking, are context-free resources for interactants to use in specific contexts where they become context-sensitive (Seedhouse 2004, p.42). The task of a CA analyst becomes to ‘explicate the structural organization of talk in interaction at this interface between context-free resources and their context-sensitive applications’ (Hutchby & Wooffitt 1998, p.360).

2) *Corpus linguistics*

The analysis in this study will be based on a corpus of transcribed classroom teaching and the IELTS speaking test for the 27 subjects. However, the qualitative analysis is going to focus only on the classroom data as the aim of this phase of the research is to answer the third sub-question on the characteristics of classroom interaction and its impact on the subjects' lexical diversity. Corpus linguistics in this sense will be very useful in providing empirical data that can help examine the subjects' language performance from an emic perspective as recommended by Seedhouse's framework of classroom interaction analysis.

Kennedy (1998, p.4) posits that a corpus 'can be analysed distributionally to show how often particular phonological, lexical, grammatical, discoursal, or pragmatic features occur, and also where they occur.' He indicates that corpus linguistics has a tendency to focus on lexis and lexical grammar rather than on syntax because it uses concordance that analyses single lines to find contextual evidence rather than longer stretches of text that can help analyse syntax or discourse. This is, indeed, what this research study intends to do in order to examine the lexical diversity of the subjects and to make inferences related to language use in the classroom.

Descriptive linguists try to make use of computerised corpora to describe in a reliable way the actual and the probable uses of lexicon and grammar of languages. It is the probabilistic aspect of this description that distinguishes corpus-based linguistic studies from conventional field-based linguistic studies.

Kennedy (1980) distinguishes between *general corpora* (or core corpora) that are gathered for unspecified linguistic research and *specialised corpora* that are designed and gathered for specific linguistic projects. *General corpora* are usually a balanced mix of spoken and written texts from different genres and domains of language use. *Specialised corpora* usually focus on regional or sociolinguistic variation including *dialect corpora*, *regional corpora*, *non-standard corpora*, and *learners' corpora*. This latter applies to the corpus gleaned for the specific purpose of this study to examine the subjects' lexical diversity and, therefore their language proficiency in the classroom.

Most of the corpus-based grammatical and lexical studies are based on the analysis of written corpora despite the fact that the spoken form of language is much more used than the written

form. Kennedy justifies this by the difficulty of transcribing spoken language because it involves ‘complex phonetic and prosodic features’. Descriptive corpus studies have approached English language from different angles including lexis, morphology, syntax and discourse. Kennedy claims that corpora are very essential nowadays for effective and comprehensive lexical descriptions.

One of the concerns in counting words based on their graphic forms is the inability to distinguish the word functions (whether a word is used as a noun, adjective, or verbs, etc.). That is why lemmatisation is used in recent corpus studies where words of the same type but with different functions (inflections) go under the same lemma or headword. Lemmatisation is indeed used in the analysis of the collected data to distinguish between the forms of content words in order to get a clear idea about the lemmas used in the classroom and in the IELTS test and the differences or similarities between them.

Transcribing spoken language without having agreed-on transcription rules creates inconsistencies in graphic word counts. However, analysts tried to reduce the impact of the absence of such rules by treating word forms that have related inflections or derivations as belonging to one family known as lexeme or lemma. Kennedy defines lemmatisation as ‘a process of classifying together all the identical or related forms of a word under a single entry’ (Kennedy 1998, p.207).

3) *Walsh’s combined approach (CLCA)*

While Seedhouse’s (2004) framework of analysis of the L2 classroom interaction will be useful to analyse transcripts of classroom teaching, and corpus linguistics will be useful to provide statistical data for qualitative analysis, Walsh’s (2011) combined approach of Corpus Linguistics and Conversation Analysis (CLCA), will provide ‘detailed, micro-analytic descriptions of spoken interaction’ (Walsh 2013, p.106). In fact, his approach will facilitate the task of conducting an analysis that combines qualitative data drawn from Seedhouse’s framework and quantitative data drawn from corpus linguistics. Frequency word lists, concordances and key words will be referred to in the qualitative analysis of some excerpts from class interaction by joining inferences based on them with inferences based on conversation analysis.

Walsh (2011) believes that conversational analysis can be applied in a classroom context to examine ways in which teachers and students create contexts in relation to their goals. The interactional pattern that is typical of a classroom is different from any other context considering the specific roles undertaken by participants in different contexts.

In a language classroom, for example, most interactions are related to the enterprise of learning a second language; turn and topic management, sequential organisation and choice of lexis are all determined by that enterprise and by the roles of interactants. (Walsh 2011, p.85)

Walsh indicates that classroom discourse is different from conversation but a number of intersecting features, like turn-taking, topic switches, false starts, hesitations, and so on make conversation analysis very relevant to the classroom context. He clarifies that conversation analysis does not use an imposed structure to make class interaction fit into it as in system-based approaches and in the functional approach, but interprets the existing data as it presents itself. In fact, what makes conversation analysis different from other approaches is that it does not use a template for all contexts but can lend itself to different contexts like a classroom situation by asking questions about teaching and learning. Walsh (2011) calls this an applied form of conversation analysis that derives understanding of class discourse from classroom goal-oriented activities and from participants' instantaneous co-construction of meaning. According to Walsh (2011) a corpus is a collection of spoken or written texts that is stored electronically and that can be searched using special software.

Walsh's (2011) suggestion to combine corpus linguistics and conversation analysis is based on his view that the relation between them is a relation of complementarity, where corpus linguistics covers large texts and disregards context, while conversation analysis examines meticulously the details of highly contextualised shorter texts. The main focus of corpus linguistics is lexis, including single words, combinations of words and word clusters, whereas the main focus of conversation analysis is on adjacency pairs, turn-taking, preference organisations, and repair.

Not only are they mutually beneficial, they actually offer each other synergies and enable a deeper, richer level of analysis. The approach is iterative: it requires a switch from CL [corpus linguistics] to CA [conversation analysis], back to CL and then on to

further CA. One approach informs the other, provides directions and enables closer analysis. (Walsh 2011, p. 103)

Walsh indicates that this combined approach has already been used by Walsh *et al.* (2011) in a study of classroom discourse on small group teaching in Irish universities. They have found that CLCA approach (Corpus Linguistics and Conversation Analysis) permitted the description of features of spoken discourse both at micro and macro levels; i.e. word and text levels (Walsh 2011, p.100).

3.10 Research validity and reliability

Teddlie and Tashakkori (2009, p.209) state that researchers using mixed methods usually face the challenge that ‘they use two different sets of standards for assessing their data quality’. One set of standards is for qualitative methods, and the other set is for quantitative methods. They define validity as the extent to which ‘data represent the constructs they were assumed to capture’, and reliability as ‘whether the data consistently and accurately represent the constructs under examination’. According to Teddlie and Tashakkori (*Ibid*)

The researcher must answer two basic questions pertaining to data quality. The first question concerns *measurement validity/ credibility*: Am I truly measuring/ recording/ capturing what I intend to, rather than something else? (...) The second question involves the *measurement reliability/ dependability* of the data: Assuming that I am measuring/ capturing what I intend to, is my measurement/ recording consistent and accurate (i.e. yields little error)?

3.10.1 Validity

In the case of the present study, measurement validity has been clearly established in both the quantitative and qualitative strands of the research. Both the mock IELTS speaking test and class recordings helped to capture what the main research question is investigating, which the subjects’ lexical diversity on the IELTS test and in the classroom. Using audio recordings of the subjects while taking the test and while teaching in the classroom facilitated the examination of the main construct, which is their speaking skill through a quantitative and qualitative analysis of their lexical diversity. The quantitative data extracted from the recordings in the form of rates

and word lists helped to measure what is intended to be measured which is the subjects' lexical diversity that is used for data interpretation at a following stage. However, one might argue that the quantitative instruments that are used to answer the second sub-question on the correlation between IELTS scores and the mentoring college-teachers' (MCT) scores might not be as solid in terms of validity as the instruments used for the first sub-question based on the argument that MCT scores do not reflect the subjects' language proficiency but their overall performance on teaching practicum as explained in section (3.8.1). This argument suggests that there are two threats to the validity of the data extracted from the MCT grades when they are used for a correlation with IELTS speaking scores:

- 1- MCT scores are not about language proficiency but overall performance on teaching practicum.
- 2- The validity of IELTS is well established and acknowledged in the available literature on language proficiency testing, whereas MCT assessment is not.

For the first concern, I argue that the MCT score represents the totality of the criteria that is considered in assessing a student teacher's potential to be a competent teacher of English after graduation. English Language proficiency is, therefore, one construct among others that is not assessed separately because it does not function separately in a teaching context but in interaction with other constructs. The competencies that the MCT score reflects do have a language component that the MCT observes and assesses as it is performed in its context. Competency number 3 which is *implementing and managing learning* (appendix 1) has a *language and delivery* component that lists the following language-focused criteria:

- Uses accurate and appropriate language, including pronunciation, stress and intonation
- Grades own language to the level of the students
- Uses classroom language accurately and fluently
- Maintains target language focus throughout the lesson

Other criteria pertaining to different competencies have a pervading English language component, such as:

- Competency 1- *Professionalism and understanding*: Gives peers constructive feedback

- Competency 2 - *Planning for learning*: prepares carefully structured activities aimed at teaching:
 - *Writing as a skill – for example process writing
 - *Writing as language reinforcement – for example, gap fill (age/level appropriate)
 - *Grammar – presents grammar through a variety of meaningful contexts
 - *Grammar – prepares activities which promote controlled and freer practice of grammar
- Competency 3 - *Implementing and managing learning – language and delivery*: uses a range of effective questioning and elicitation techniques.
- Competency 3 - *Implementing and managing learning – classroom management*: establishes and maintains rules and clear routines for behaviour in the classroom.
- Competency 3 - *Implementing and managing learning - Communication skills*: builds rapport through interaction with students
- Competency 4 - *Monitoring and assessment*: marks and monitors learners’ work providing constructive oral and written feedback
- Competency 5 – *Reflection*: discusses and justifies own pedagogical decisions

Therefore, the use of the MCT’s score for a comparison with the IELTS score is validated by the fact that it assesses the student teachers’ language proficiency as it interacts with the context where it is meant to be used. The IELTS score is used to assess students’ language proficiency for a potential use in the same context where the MCT score is assigned. The data drawn from the correlation of the two sets of scores is therefore valid to tell if IELTS scores can predict scores assigned in the teaching context where English language proficiency is performed.

For the second concern regarding the comparison between IELTS as a valid international test and MCT assessment which has no records of validation, I argue that MCT assessment gained its validity from its history as it has been used across the six colleges in UAE where the Education program is run since its inception in the year 2000. No concerns or issues have been raised since then regarding its validity despite the fact that many faculty and program chairs who have been using it have high qualifications in education and many of them are researchers in the field. So, MCT assessment and MCT scores gained ‘institutional’ validity over the last

sixteen years that qualified them to be benchmarked against IELTS scores in the present study. Beside this internally developed validity, MCT assessment gains its validity from an external accreditation of the program from the University of Melbourne, Australia. In fact, the program was first developed by experts from the Education department at the University of Melbourne who worked collaboratively with specialists in education at HCT (Higher Colleges of Technology, UAE). All the program courses and assessments were designed, reviewed and approved by both HCT and the Education department of Melbourne University. Auditors from the University of Melbourne continue to visit the six colleges every two years to review the program policies and practices before renewing the accreditation. The program has always received positive reports from the auditors and recommendations for improvement have never addressed the Teaching Practicum (TP) or its assessments as it has been considered as one of its strongest components.

The validity of the qualitative strand of the research is well established through an examination of the subjects' recordings in the classroom while taking an emic perspective. Examples of the subjects' speech acts were used to provide detailed analysis of the impact of classroom interaction on the subjects' lexical diversity. Inferences based on the analysed samples of the subjects' speech in the classroom provided valid interpretations of the effect of classroom interaction on the subjects' lexical diversity.

3.10.2 Reliability

Though the validity of the research has not been a real concern due to the arguments that have been discussed in the previous section, reliability represented a challenge due to some uncontrollable factors. First, using a mock IELTS test instead of a real one is a forced choice since it was not possible to get the recordings of the real test that the participants had taken right before this research study started. Cambridge ESOL was contacted to provide copies of the subjects' real speaking test, and was informed of the research purposes but a negative response was received. Thus, the first threat to the reliability of the data collected from the 'mock' test is that it had to use some published tests by Cambridge ESOL or by authors who specialised in IELTS preparation course books. The researcher resorted to the librarian of the college who helped to choose good but less used resources.

In the administration of the test, the challenge was to make sure the mock test is administered and taken by the candidates in the same conditions as the real test. As a certified examiner of the IELTS test, the researcher tried to replicate the same test conditions in order to maximise reliability. As a matter of fact, the researcher did not conduct the test by himself but asked five colleagues who were certified IELTS examiners and who were not teaching in the Education program to conduct the mock test. The subjects were divided into four groups that took different tests on different days due to logistic reasons related to the availability of the examiners and the subjects. Subjects taking the same test were examined on the same day and special measures were taken not to allow the ones who took the test first to disclose the content of the test to their mates in the same group.

Another source of threat to the reliability of the test was the fact that the subjects knew it was not a real one, so there was a risk that some of them might not take it seriously. To minimize the impact of this threat, the researcher met with the subjects before taking the test to raise their awareness of the importance of the test for the research and the possible impact of the research findings on the future of the program graduation requirements. As for the quality of the recordings, the researcher provided quality devices and materials for the IELTS examiners and for the subjects to record their classes to make sure the collected data is clear for transcription and analysis.

3.11 Research Ethics

Ethical considerations in relation with the well-being of the participants in this research study have been taken into account and documented from the beginning of the process. A *Research Proposal and Ethics Clearance Form* (appendix 135) was submitted to HCT research committee. The form provided all the necessary information about the research including details of the researcher, the research title, a brief description of the research, intended participants, research procedures, benefits and risks of the research for the participants and the institution. A detailed research proposal was also attached to the form. Upon receiving the approval by HCT Research Board and following the sociocultural norms in UAE, parent/ guardian consent forms were issued to potential participants both in English (appendix 136) and in Arabic (appendix

137) to ensure that they give permission for their dependents to participate in the research study. The consent forms stated the purpose of the research study, and included the researcher's commitment to anonymity by protecting the identity of the participants, and to codes of confidentiality by keeping the collected data in a private and safe place.

3.12 Methodological issues

Despite different precautions and measures taken before implementing the study in its quantitative phase then its qualitative phase, some issues emerged at different stages. Some were easy to control so they do not have any significant impact, but others were difficult to keep under control and could have impacted the collected data and the subsequent analysis even at a low significance level. Some of those issues relate to the sample size which was not determined by the researcher nor by the student teachers who wanted to participate in the study but by their parents and guardians whose consent limited the sample to 27. Had it not been the parents' decision, the sample could have reached at least 40 participants which could have added value to the meaningfulness of the findings and the drawn conclusions.

Another challenge was to manage to convince the subjects to treat the mock IELTS test as if it were the real one. In reality, it was not possible to monitor this fact to make sure the subjects would perform at their best level, especially for those who had taken the test and achieved the required band.

It was also not possible to ensure the examiners follow IELTS procedures in the same way they would in the real test. As a matter of fact, the duration of the mock test varied considerably between 9 and 15 minutes though the time range that is recommended by IELTS regulations is between 11 and 14 minutes. The reasons are presumably due to the fact that some examiners did not abide by the official IELTS instructions from the beginning to the end of the test, especially with regard to formulaic rubrics that the examiners are supposed to read. It is also possible that the examiners did not keep watching the timer for each part. By listening to the five recordings that lasted less than 11 minutes, it is clear that some examiners avoided the formulaic structures of the test especially in parts 1 and 2 and used their own words to make it shorter and simpler for themselves and for the candidates. It is also clear that those examiners

ignored the optional follow up question ‘why’ after the candidate answers personal questions especially in parts 1 and 2. For those who went above the maximum time, it is noticed that the examiners were flexible in granting more time for the candidates who showed willingness to produce longer stretches of speech, especially in parts 2 and 3.

For the class recordings, I refrained from attending classes and recording them by myself to avoid the ‘observer effect’ on the performance of the subjects, and I preferred to ask the them to record their own classes and submit them to me. However, I found that 4 recordings were noticeably below the minimum time for a primary class which is usually around 35 minutes. The shortest one lasted only for 11 minutes and 33 seconds, another one lasted for 13 minutes and 34 seconds, a third one lasted for almost 15 minutes, and the last one lasted for 21 minutes and 32 seconds. There were no clear reasons why the recordings were discontinued but the assumption is that some unknown technical problem that happened while recording without being noticed by the student teachers as they were busy teaching. Nevertheless, those recordings were not discarded from the analysis as they could still reflect the students’ speaking performance in class.

Transcribing class recordings was a challenging task due to the complexity of classroom interaction in which turn-taking mechanisms with their turn relevance place are not always taken into consideration, especially when school students get excited about an activity or a question asked by the teacher, and when the teacher has difficulties with classroom management. School students raise their voices while the teacher is speaking which makes it difficult to transcribe some parts.

3.13 Conclusion

In this chapter, I have clarified my research paradigm and related epistemological background. I have also presented my methodological approach to research followed by my research design for the present study and my data collection and data analysis procedures.

Reflecting on the research as a process from the moment the main question started to develop through the methodological assumption and procedures, to the final findings and

recommendations, I believe that the path of inquiry is multi-dimensional in a sense that it cannot be explored from just one standpoint. In this research study, a pragmatic approach to developing knowledge is adapted to the context where the study is conducted. Pragmatism helped to use different paradigms and different methods in a way that they complemented each other to help achieve clear and detailed answers for the research questions. The answers obtained answer satisfactorily the main question. As a researcher, I think using different approaches with regard to the researched knowledge, makes inquiry a richer process with richer outcomes.

Chapter 4: Results

4.1 Introduction

In this chapter, I am presenting quantitative and qualitative analyses of my data. The quantitative part (4.2) will present the results and the analysis of the correlations between IELTS and class lexical diversity scores (D-scores), and between Mentoring College Teachers' (MCT) grades and IELTS scores. The second section (4.3) will present a corpus linguistics analysis of content-word frequency lists for both the IELTS test and the classroom teaching. A comparison between content-word frequencies in each context will be conducted to draw conclusions on the differences between the two contexts. The qualitative findings will be presented in section (4.4) with an analysis of chosen sequences of the subjects' talk in the classroom to focus on classroom interaction and its impact on the lexical diversity of the subjects.

4.2 The quantitative analysis

In this first part, I will present the subjects' lexical diversity scores in the IELTS speaking test and in class as computed by VOC-D. Then, I will present the findings of two correlations: (1) between the subjects' D-scores on IELTS and in class, and (2) between MCT grades and IELTS scores. I am also presenting the findings of corpus linguistics analysis of the subjects' talk in the IELTS test and in the class.

4.2.1 The correlation between IELTS and class D-scores

The subjects' lexical diversity scores on the IELTS speaking test and in the classroom have been computed by VOC-D which generated D-scores illustrated in table (4.1). As noted in chapter 3 (sub-section 3.6.2), VOC-D is a mathematical formula of calculating vocabulary diversity that overcomes issues of text-length effect associated with previously used TTR measure. McKee, Malvern, and Richards (2000, p.323) present VOC-D as a software that 'automates' vocabulary measurement from texts that are prepared for processing by CHAT and CLAN of the CHILDES project (see sub-section 3.6.1). What characterizes VOC-D is that it does not calculate lexical diversity in a linear way as TTR does to avoid the effect of the length

of the text on the ratio by increasing the number of tokens and decreasing the number of types. Conversely, VOC-D relies on calculating the probability of new words in longer texts through computing the value of 100 trials of a random sample of 35 tokens (without replacement). Then, the same procedure is repeated for random samples of 36 to 50 tokens to create an empirical TTR curve that is based on the mean of each 100 sample. That curve is plotted against a theoretical curve which is created based on a formula using \mathcal{D} coefficient (Malvern et al. 2004, p.51) to find the best fit between them, which is the value of D . According to McCarthy and Jarvis (2010, p.383) D values ‘tend to range between 10 and 100’.

Table 4.1 *Student teachers’ D-scores for both IELTS and class*

	Students	D-score on IELTS	D-score in class
1	T01	86.13	58.16
2	T02	78.63	65.67
3	T03	53.09	71.46
4	T04	79.16	31.70
5	T05	63.24	42.25
6	T06	73.22	57.28
7	T07	70.21	65.67
8	T08	72.77	75.95
9	T09	69.58	58.49
10	T10	61.66	59.54
11	T11	69.99	56.66
12	T12	63.67	32.55
13	T13	64.55	63.17
14	T14	50.45	74.07
15	T15	72.07	58.98
16	T16	68.56	57.81
17	T17	76.15	29.96
18	T18	64.26	56.66
19	T19	70.66	42.52
20	T20	64.88	43.74
21	T21	77.55	54.81
22	T22	62.59	46.20
23	T23	73.96	60.37
24	T24	62.29	50.67
25	T25	68.04	70.34
26	T26	76.50	65.51
27	T27	70.55	71.65

The figures displayed in table (4.1) show clear discrepancies between the performance of the students on the speaking component of IELTS test and their performance in a class teaching situation. A total of 22 out of 27 student teachers (i.e., 81%) displayed higher levels of lexical diversity on IELTS than in class situation (highlighted in grey colour in table 4.1). However, the differences between IELTS D-scores and class D-scores for those 22 students vary considerably from 1.38 (T13) to 47.46 (T04). These differences may be entirely ideolectal, but it may also be related to the types of questions which formed part of the IELTS test. For instance, it was reported, unsurprisingly, by some students that they were more confident answering questions on topics that they were familiar with. For instance, some subjects had questions about shopping, films, and social events (test 2-appendix 3 and test 3-appendix 4) which could have contributed to raising their IELTS D-scores because they were in possession of a greater range of vocabulary to use in conversations about these topics. On the other hand, there were subjects who had questions about football matches, football celebrities and superstition matters (test 1-appendix 2 and test 4 - appendix 5), with which they were less acquainted, which would have inevitably had an effect on the IELTS D-scores (i.e., it would have reduced their lexical diversity). Where student teachers had lower D-scores in IELTS, the resulting difference between these and the classroom D-scores would have been lower, too. The difference between class and IELTS D-scores may also be explained by contextual variation in the classroom situation. For instance, teaching young children may result in a still lower class D-score compared to teaching older children. So, the discrepancy between the IELTS and class lexical diversity needs to consider the contextual variables in both settings in order to account for it.

On a lexical diversity scale of 100 (McCarthy & Jarvis, 2010), an arbitrarily-chosen 10-point difference between the two performances could arguably be considered as a clear difference based on which inferences can be made. A further examination of figures in table 4.1 reveals that 18 students out of the 22 who had higher D-scores on the IELTS test exceeded the 10-point difference. They represent almost 82% of the total D-scores that are higher on the IELTS test than in class. This majority illustrates clear superiority of lexical diversity in the IELTS context compared with the class context for more than 80% of the students.

Conversely, five subjects (T03, T08, T14, T25, T27) had lower D-scores on IELTS than in class. Though they represent only 18.5% of the total population, important questions arise regarding

the reasons why they did not follow the majority, and if this has any implications on the conclusions drawn from this analysis. Before trying to answer those questions, it is worth noting that only two subjects have clear superior D-scores in the class compared to IELTS, namely T03 and T14. The other three subjects have very little differences. Based on my knowledge of the context of the study and the participants, who are my students, I can presume that some personal factors interfered in making IELTS D-scores clearly lower than class scores. The first factor that I find plausible is that they did not take the IELTS test seriously because they had no genuine interest in taking it, especially that some of them took it after they had taken the official one and after achieving the required band 6 score. This applies especially to T14 who was already a year 4 student at the time she took the mock test, and T03 who was a year 3 student who had passed the test. The second factor that can possibly justify this irregularity compared with the majority of the participants is the difficulty they faced in answering questions about sports and sport celebrities in parts 2 and 3 of the test, and this applies again to T03 and T14. By revisiting their interviews, I noticed that they relied a lot on repetition of the same lexical items as an indication of the limitation of their knowledge of that topic. A third possible reason could be psychological, which relates to the fear of the test and the strict test conditions. I noticed some of this factor especially by listening to T14 who displayed a lot of hesitation while answering the questions. However, all this reasoning to justify the irregular pattern displayed by 5 subjects remains hypothetical until a proper research study is conducted to look in depth into the possible reasons and their relationship with language performance on the test.

1) Test of normality

The validity of any statistical procedures and related conclusions depends on the normal distribution of the data. Statistical procedures like correlations, regression, t-test, and so on are based on the assumption that the analysed data is normally distributed. Ghasemi & Zahediasl (2012, p.486) assert that ‘normality and other assumptions should be taken seriously, for when these assumptions do not hold, it is impossible to draw accurate and reliable conclusions about reality’. Moreover, Ghasemi and Zahediasl (2012) indicate that for sample sizes equalling 30 or more normality should not represent an issue, but for less than 30 a test of normality should be run to check that data is normally distributed. In the case of the present study, the sample size is 27, and the intention is to compute 2 correlations. Therefore, running a test of normality is very important for the validity of the analysis and the drawn conclusions. Two correlational tests

are run in this research study for which tests of normality are conducted: (1) the correlation between D-scores in IELTS and in class, and (2) the correlation between IELTS scores and MCT grades.

Prior to the computation of the correlation between D-scores of IELTS and class performances, Shapiro Wilk test is conducted as a correlation and regression test of normality (Yap & Sim 2011, p.2143) to verify that D-values were normally distributed. The assessment revealed that both sets are normally distributed with $p > 0.05$, as shown in Table (4.2).

Table 4.2 *Test of normality for the IELTS and class D-scores*

	Shapiro-Wilk		
	Statistic	Df	Sig.
IELTS	.935	27	.092
Class	.954	27	.263

2) The mean

In line with the findings discussed above, SPSS computation of the mean values of IELTS and class Ds for the whole group of subjects (appendix 6) confirms that the IELTS context results in a higher D-score, and the difference between the two means is about 12 points (69.05 on IELTS to 56.36 in class). This confirms what we saw with individual students just previously that there is a clear difference in lexical diversity between the performances of the subjects in those two different contexts.

3) The standard deviation

The standard deviation figures (appendix 6) show that there is more variability of D-scores when it comes to class performance (12.8) compared to the test situation (7.85). It might be suggested that this arises because of the differing contexts involved. The test situation is relatively homogeneous, and students are usually well prepared for it. This is not the case for the classroom situation where there is variation in who is taught and what they are taught. Student teachers in this study gave classes to learners ranging from grade 1 to grade 8, which suggests a wide range of topics and of language complexity. Given that, performance is more likely to vary in the classroom situation compared to the test situation.

4) Pearson correlation for D-scores of IELTS and class performances

The relationship between student teachers’ D-scores on IELTS test and in the classroom was investigated using Pearson product-moment correlation coefficient. As noted above, preliminary analyses were performed to ensure that there were no violation of the assumptions of normality and linearity. SPSS results show that there is no correlation between the two variables with

$r = -.160$, $n=27$ and a non-significant relationship with $p>.05$ (2-tailed) as illustrated in table (4.3)

Table 4.3 *Pearson correlation of IELTS and class D-scores*

		Students’ D on IELTS	Students’ D in Class
Students’ D on IELTS	Pearson Correlation	1	-.160
	Sig. (2-tailed)		.425
	N	27	27
Students’ D in CLASS	Pearson Correlation	-.160	1
	Sig. (2-tailed)	.425	
	N	27	27

The statistical findings mentioned above give evidence that the performance of the subjects on IELTS and in class are clearly different which is reflected in differences in the D-scores and the Pearson correlation that indicates that there is no relationship between the two sets of scores.

4.2.2 The correlation between MCT grades and IELTS scores

IELTS test scores as assigned by the examiners who conducted the mock speaking test have been correlated with the MCT grades assigned to the subjects on TP placement in schools. MCT grades have been converted from a scale of 100 to a scale of 9 -like IELTS scale- to make numerical comparisons for individual students easier (see conversion table in appendix 8) . The scores are illustrated in table (4.4) below:

Table 4.4 *MCT grades and IELTS speaking test scores*

	Students	MCT grades	IELTS Scores
1	T01	8.5	6.5
2	T02	7	5.5
3	T03	6.5	6
4	T04	6.5	6.5
5	T05	7	5
6	T06	7.5	5.5
7	T07	6.5	5.5
8	T08	7.5	6.5
9	T09	7	5.5
10	T10	7.5	5.5
11	T11	8	5
12	T12	8	5.5
13	T13	5.5	6
14	T14	7.5	5.5
15	T15	6.5	5.5
16	T16	6.5	6.5
17	T17	6.5	6
18	T18	7	6
19	T19	8.5	6.5
20	T20	7	5.5
21	T21	8.5	5.5
22	T22	8	5.5
23	T23	7	5.5
24	T24	7	5.5
25	T25	8	5.5
26	T26	8	6
27	T27	6.5	7

The test of normality for the correlation between MCT grades and IELTS scores shows that scores are normally distributed with $p > .05$ as shown in table (4.5):

Table 4.5 *The test of normality for the MCT and IELTS scores*

	Shapiro-Wilk		
	Statistic	Df	Sig.
Students' D in CLASS	.939	27	.116
MCT grades	.928	27	.063

Pearson correlation revealed the following statistics in table (4.6):

Table 4.6 *Pearson correlation of MCT grades and IELTS speaking scores*

		MCT grades	IELTS Scores
MCT grades	Pearson Correlation	1	-.142
	Sig. (2-tailed)		.479
	N	27	27
IELTS Scores	Pearson Correlation	-.142	1
	Sig. (2-tailed)	.479	
	N	27	27

SPSS results show that there is no correlation between the two variables with $r = -.142$, $n=27$ and a non-significant relationship with $p>.05$ (2-tailed) as illustrated in table (4.6)

This finding is in total agreement with the correlation of the D-scores, which reinforces the argument that there is no relationship between subjects' performance on the IELTS speaking test and their performance in the classroom.

4.3 Corpus linguistics analysis

To account for the quantitative discrepancies between the IELTS test and the class discussed in (4.2), this section will present a corpus linguistics analysis using Wordsmith tool that provides *content-word frequency lists and their concordances to disambiguate word classes for both the IELTS speaking test and the class. A quantitative comparison between the top-twenty content words on each list will be presented first, then, a detailed analysis of the class top-twenty words and their lemmas will be used in a qualitative analysis of classroom interaction following Walsh's CLCA approach and Seedouse's framework for the analysis of classroom interaction (see 3.7.1 and 3.7.3).

* Content words 'convey the meaning of a sentence and correspond to open-class words such as nouns, verbs, adverbs, and adjectives'. (Chamberland et al. 2013, p.94). However, function words play a grammatical role and correspond to closed-class words like articles, prepositions, and pronouns.

4.3.1 A comparison between IELTS and class content-word frequency lists

The statistics of the top twenty content-word frequency list are summarised in table (4.7)

Table 4.7 A summary of the rates of the top twenty content-words in IELTS and in the class

	IELTS	CLASS
Verb	10/20 = 50%	14/20 = 70%
Noun	6/20 = 30%	04/20 = 20%
Adjective	02/20 = 10%	02/20 = 10%
Adverb	02/20 = 10%	-

A graphical representation (figure 4.1) of table (4.7) may serve as a useful tool to visualize the different proportions of content words in the IELTS speaking test and the class.

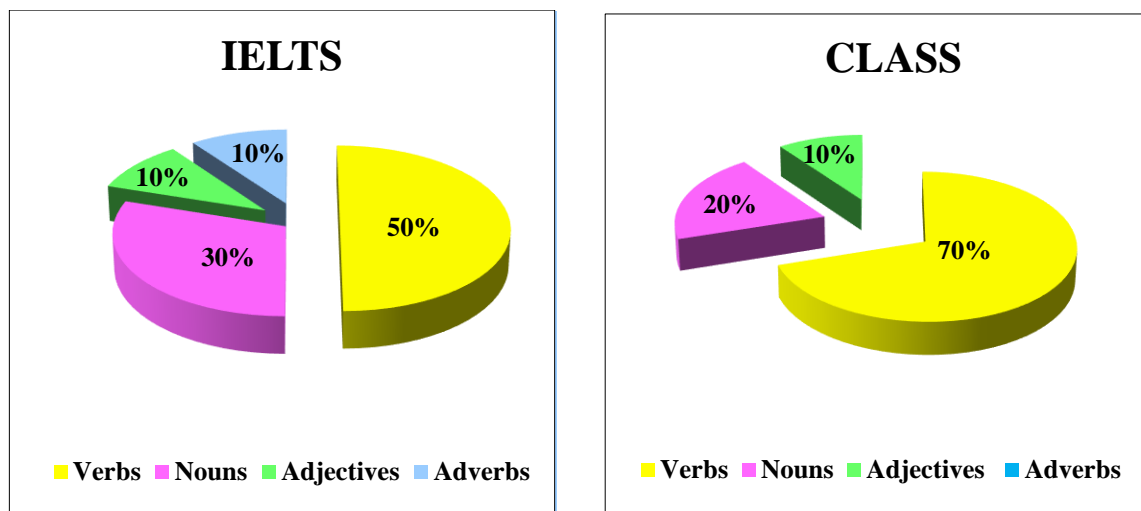


Figure 4.1 Proportions of content words in IELTS and in class

The table and the pie chart illustrate the dominance of verbs in both contexts. However, the class top twenty content word list is largely dominated by verbs (almost three quarters) that pertain to simple classroom jargon like *write*, *look*, *listen*, *sit*, and so on (Table 4.8). They are repeated at a high rate either for academic purposes to develop students' literacy skills as in 'just *write* what you can see' or for classroom management as in '*look* here, Layla and Fatmah *look* here'.

In fact, continuous repetition of the same instruction verbs like *listen* to draw students' attention or *read* to carry out class or group tasks pushes down the lexical diversity values.

High-rate repetition of simple action verbs is, then, an essential feature of class language that deem the lexical diversity to be at a lower level compared to IELTS language which, on the contrary, relies less on verbs and a little bit more on nouns, adjectives and adverbs to discuss and support one's arguments vis-à-vis contemporary topics and issues.

4.3.2 An analysis of class content-word frequency list

Along with individual word lists for each subject (appendices 10-36), a generic content-word frequency list illustrating the first twenty most frequent words (expressed in raw numbers) for all the subjects in the IELTS speaking test and in class was generated (appendix 9). The top twenty words on the generic class list show that 14 words are verbs (i.e., 70%), and 4 are nouns (i.e. 20%). As it shows in table (4.8), all the verbs are clearly associated with the context of the classroom in which they are produced. It is worth noting here, that three verbs, namely *be*, *do*, and *have* are on the top of both the class and the IELTS lists due to their dual role as verbs (content words) and auxiliaries (function words). If we exclude the three 'semi-content' words on the top of the list (i.e., *be*, *do*, and *have*) the remaining 11 verbs either express classroom instruction related to academic skills that characterize the class context, like *write*, *read*, *listen*, or action verbs associated with carrying out class activities like *go*, *come*, *sit*, *look*, *see* and *finish*. The last verb on the list, which is *know* is a perception verb that is clearly associated with the teaching and learning context as well. The remaining 6 words on the list comprise 4 nouns (*group*, *number*, *girls* and *sentence*) and 2 adjectives (*excellent* and *good*) that represent an integral part of any learning and teaching context.

Table 4.8 List of verbs in the top-twenty word-list of the class

	Verb	Frequency	Rank
1	Be	1742	1
2	Do	786	2
3	Have	505	3
4	Write	281	6
5	Go	269	7
6	Come	251	8
7	Sit	245	9
8	Read	244	10

9	Want	230	11
10	Listen	202	13
11	Look	186	15
12	See	170	16
13	Finish	156	18
14	Know	146	19

1) Verb lemmas

- Lemma *Write*

Lemma *write* occurs 281 times in the generic class word frequency list (appendix 9) with the verb form *write* occurring 251 times, *writing* occurring 19 times, *written* 5 times, and *wrote* 6 times. A detailed examination of the concordances of the most frequent verb form *write* reveals that it is mostly used for direct class instructions (173 occurrences) when the teacher instructs her students to practise writing, such as ‘just *write* what you can see’ and ‘*write* it on the map’, while the remaining uses (78 occurrences) are for the act of writing as in ‘I will *write* it on the board’ and ‘do you know how to *write* these sentences?’ (appendix 37).

- Lemma *go*

Lemma *go* occurs 269 times. The form *going* counts 152 times, *go* counts 106, and *went* counts 11 times (appendix 9). By looking at the collocations of the verb form *going* (appendix 38), we notice that the highest collocation is with the preposition *to* (134 times) to make up the future form *going to*, which is used by teachers to mention planned instruction such as ‘you are *going to* think of a magic invention’ and ‘we are *going to* read a story’. This repeated pattern reflects the nature of the teacher’s talk in classroom which is based on announcing and leading planned class activities. On few occasions (15 occurrences) *going* is preceded by the auxiliaries *is* or *are* to form the present progressive indicating near future action as in ‘we *are going* on a bear hunt’ and ‘*is* she *going* alone?’ The second in frequency of occurrences is the verb *go* which is divided between direct instruction for class activities such as ‘*go* and find me letter ‘f’ and ‘*go* back to your seats’, and talking about actions like ‘they cannot *go* over it’ and ‘where did you *go*?’ (appendix 38).

- Lemma *come*

Lemma *come* occurs 251 times. The verb form *come* occurs 227 times, *came* occurs 12 times, *coming* occurs 8 times and *comes* occurs 4 times (appendix 9). The verb form *come* is used mainly to give some students instructions (164 occurrences) to move to the front of the class to perform some actions like ‘Shahad can you *come* here?’ and ‘Yusuf *come* and write number 19’. ‘*Come on*’ is also used by teachers (51 occurrences) to encourage their students to take part in class activities as in ‘*come on*, let’s go there’ and ‘*come on* jump! jump! jump!’. On 12 occasions *come* is used as an action verb as in ‘sit down, I will *come*’. (appendix 39)

- Lemma *sit*

Lemma *sit* occurs 245 times. In all occurrences, *sit* is used to instruct students to sit down whether at the beginning of the class when students greet their teachers while they are standing as in ‘I’m fine thank you, *sit* down please’ or during class activities as in ‘*sit* down Fatmah, I am coming’. (appendix 40)

- Lemma *read*

Lemma *read* occurs 244 times with the verb form *read* occurring 229 times and *reading* occurring 15 times. By examining the concordances of the verb *read* (appendix 41), we notice that it is used 220 times to give instructions such as ‘who will *read* the title?’ and ‘*read* the answer’ and only 9 times to refer to the act of reading like ‘I am going to *read* first’ and ‘do you want to *read* it?’. The form *reading* occurs 13 times as a gerund to refer to the act of reading as in ‘who will complete *reading*?’, and 2 times only as the present progressive form of *read* as in ‘you are *reading* about movies’ (appendix 41).

- Lemma *want*

Lemma *want* occurs 230 times with the verb *want* occurring 201 times and *wants* occurring 29 times. By looking at the concordances of the verb *want/wants* (appendix 42), it is noticeable that it is mainly used for two purposes, either to ask students if they wanted to engage in some activities (120 occurrences) as in ‘do you *want* to read it?’ and ‘who *wants* to act?’ or to give instructions related to class activities (75 occurrences) as in ‘I *want* you to write a paragraph’ and ‘I *want* you to work in pairs’. On the remaining occasions, teachers used *want* in affirmative

sentences as in ‘show your friends what your group *want* (sic) to invent’ and ‘he *wants* a beef burger’ (appendix 42).

- Lemma *listen*

Lemma *listen* occurs 202 times with the verb form *listen* occurring 201 times and *listening* just one time. The concordances of the verb *listen* (appendix 43) show that 198 occurrences are identified as giving instructions. Some of those instructions are given to make the students practice the listening skill as in ‘you will *listen* first then you will sing’ and ‘*listen* to this story’ but some others are used for classroom management whereby the teacher uses the verb *listen* to draw students’ attention to her when they are busy working on some tasks or when some of them are distracted by something else as in ‘*listen* to your teacher’ and ‘*listen* to me students in the back’. It is also used to engage the students in class sharing activities as in ‘*listen* to your friend’ and ‘*listen* to Shareifah’ (appendix 43).

- Lemma *look*

Lemma *look* occurs 186 times including 184 times as *look*, one time as *looks* and one time as *looking*. Most of the uses of *look* are as a verb of instruction (179 occurrences) to draw students’ attention and to order them to focus on the teacher, on one of their classmates or on a visual aid displayed for them. The collocation of the verb *look* with the preposition *at* (appendix 44) is the highest in frequency (110 times) which indicates that it was mainly used to draw students’ attention as in ‘*look* at the board’ and ‘*look* at me now and listen carefully’. On four occasions only ‘look’ is used as a noun as in ‘have a *look* first’ though it is still expressing instruction (appendix 44).

- Lemma *see*

Lemma *see* occurs 170 times including 144 times as *see* and 26 times as *saw*. The examination of the concordances of the verb *see* (appendix 45) shows that it is mostly preceded by the modal *can* to check if the students can see some displayed material as in ‘what animals can you *see*?’ and ‘can you *see* it all of you?’ The collocations of *see* as revealed by the concordance tool (appendix 45) reveals that it is preceded by *can* in 42 occurrences and by *let’s* in 39 occurrences.

Let's is used to initiate an action that involves the students such as ‘let’s *see* who will finish’ and ‘let’s *see* if you remember the letters’ (appendix 45).

- Lemma *finish*

Lemma *finish* is used 156 times with the past participle verb form *finished* occurring 102 times and the base form *finish* 54 times. Both forms are used by the teachers either to ask the students if they finished doing some assigned tasks as in ‘have you *finished*?’ and ‘did you *finish*?’, or to encourage the students to finish a task quickly as in ‘we will see who will *finish* first’ and ‘let’s *finish* it now’ (appendix 46).

- Lemma *know*

Lemma *know* occurs 146 times with the verb *know* happening 139 times, third person verb form *knows* 5 times, and the Past Simple verb form *knew* 2 times. In its three forms, lemma *know* is used as a perception verb mainly to ask students if they had previous knowledge of some concepts, vocabulary meaning, or a skill (113 times) as in ‘Children do you *know* the meaning of floor?’ and “do you *know* how to write it?”. However, sometimes ‘*know*’ is used in affirmative or negative statements (33 times) as in ‘now we *know* about the family’ and ‘I don’t *know*’ (appendix 47).

2) Noun lemmas

Only 4 words of the class top-twenty frequency list are nouns. They are listed in table (4.8)

Table 4.9 *List of nouns in the top-twenty word list of the class*

	Noun	Frequency	Rank
1	Group	207	12
2	Number	199	14
3	Girls	177	17
4	Sentence	141	20

- Lemma *group*

Lemma *group* counts 207 occurrences including 188 times in singular form and 19 times in plural form (appendix 9). Most of classes use group work, which explains why teachers address students in groups rather than individuals as in ‘this *group* has finished’ and ‘you are going to work in *groups*’ (appendix 48).

- Lemma *number*

Lemma *number* counts 199 occurrences including 194 times in singular form and 5 times in plural form. It is a high frequency word because teachers assign numbers to tasks and exercises as in ‘read this box *number* one’ and ‘now exercise *number* three’. (appendix 49).

- Lemma *girl*

Lemma ‘girl’ occurs 177 times including 166 times in plural form and 11 times in singular form. It is noticeably frequent because teachers use it to address their students who are mostly female students as UAE schools are single-sex schools and most of the subjects were teaching in schools for girls. Teachers use it to draw students’ attention or to address the whole class as in ‘girls girls girls girls listen’ and in ‘OK girls, do you like drawing?’ (appendix 50).

- Lemma *sentence*

Lemma *sentence* occurs 141 times including 115 times in the singular form and 26 times in the plural form. It is one of the top frequent words as teachers in primary level especially focus on the sentence level and try to develop students’ English language skills through studying and forming sentences as in ‘who will read the next *sentence*?’ and ‘I want someone to give me a *sentence*’ (appendix 51).

3) *Adjective lemmas*

The remaining two words in the class list of the most frequent content-words are adjectives *excellent* and *good*. What is noticeable is that they have the highest number of occurrences compared with all the verbs and nouns analysed above. *Excellent* is ranked the fourth after the three semi-content words (*be*, *do* and *have*), and *good* is ranked the fifth (appendix 9).

- Lemma *excellent*

Lemma *excellent* appears on the top of the class most frequent word-list (after the three first semi-content verbs: *be*, *do*, and *have*) with 360 occurrences which is more than any verb or noun. Teachers keep using it as feedback on students' participation in class activities and as a classroom management technique to keep them motivated throughout the lesson as in 'grey colour, yes *excellent* girls' and in 'yes *excellent* Maryam, the blue sea is deep' (appendix 52).

- Lemma *good*

Lemma *good* is the second on the class list after *excellent* and counts 335 occurrences. The form *good* counts 316, *best* counts 16 times and *better* 3 times. The form *good* is mostly used to qualify students' answers and is usually used with the intensifier *very* as shown in the collocation counts (241 times preceded by *very*) as in 'Excellent Hassan *very good*' and in 'a piece of cheese, *very good*'. It was, also used 25 times in greeting as in '*good* morning girls' (appendix 53).

4.4 The qualitative analysis

This section on the qualitative analysis of my classroom data is going to focus on the subjects' talk in the classroom in order to identify features of classroom interaction that account for the lower lexical diversity values compared with the IELTS speaking test as revealed by the statistical analysis.

4.4.1 A taxonomy of teacher repeats

A close examination of the subjects' class performance reveals an outstanding phenomenon that characterizes classroom interaction which is the considerable number of repetitions that teachers perform of their own turns and their students' turns for purposes that are going to be delineated and supported by examples in this section. In fact, repetition is a distinctive feature of the L2 classroom interaction that distinguishes it from other institutional and social interactions. It has been argued to play a central role in L2 learning (Chaudron, 1988; Gass et al., 1998; Rydland & Aukurst, 2005; Park 2014). It has different functions such as confirming correct answers, initiating repair, clarifying statements, model pronunciation, and many other functions

Table 4.10 *A taxonomy of repeat categories and types in the classroom*

	Repeat category	Repeat type	Repeated lexical items
1	Interaction-control repeats	In IRF exchanges	Lexical items that the teacher uses in an initiation turn (I) for which she does not get an immediate or an appropriate response (R)
		In group work	Lexical items that the teacher uses to monitor interaction in group-work activities
		In repair phases	Lexical items that the teacher uses in a repair phase as a way to insist on the student who made the error to correct it by herself
2	Question repeats	Elicitation & modification repeats	Lexical items used in elicitation questions or in modified forms of those questions
		Strict question-repeats	Lexical items in questions that teachers repeat a number of times with no modification
		Think-time repeats	Lexical items in questions that teachers ask twice or more to give think-time to their students
3	Feedback repeats	Confirmation repeats	Lexical items in teachers' repeats of students' answers/ contributions as a way of confirming their correctness
		Praise-word repeats	Lexical items that are used by teachers to praise students for their answers/contributions
		Repair repeats	Lexical items that teachers use to correct their students' answers/ contributions
4	Key-word repeats	Lesson key-word repeats	Lexical items that relate to the main focus of the lesson, and that the teacher uses throughout the class as all or most of the interaction with the students revolves around them.
		Activity key-word repeats	Lexical items that relate to a specific activity, but not necessarily to the lesson focus, are repeated frequently while the activity is being conducted
		Story key-word repeats	Lexical items that are part of rhyming lines in a story
5	Approach-related repeats	Language-drill repeats	Lexical items that are used in teacher turns to model the pronunciation/ spelling of new vocabulary and/ or target grammar structures
		Scaffolding repeats	Lexical items that are used in teachers' scaffolding turns
6	Procedural repeats	Instruction-clarification repeats	Lexical items that the teacher uses to clarify the procedure of carrying out an activity in a successful way, or lexical items that are used to clarify instructions for an activity when students show that they do not understand the procedure clearly
		Classroom management repeats	Lexical items that are used by the teacher to control students' attention, movement/ behaviour, and to start and finish activities.

(Chaudron, 1988; O'Connor and Michaels, 1993; Duff, 2000; Hellerman 2003; Rydland & Aukurst, 2005).

What makes those repetitions important for this analysis is their impact on the lexical diversity values of the subjects. Repetitions that use the same lexical items multiply the number of tokens at the expense of types, which as a result causes the lexical diversity values to fall. A thorough examination of the 27 classroom transcripts and their recordings revealed three realities: (1) repetition is a pivotal pedagogical factor that pervades teacher talk in all the examined classes, (2) it impacts negatively upon the lexical diversity due to its abundant use for different purposes, and (3) there are 16 different types of repetitions that serve different pedagogical purposes. They are outlined in a taxonomy of repeat categories and types illustrated in table (4.10) on the next page, and analysed in detail with specific examples from the subjects' class transcripts in the section that follows.

4.4.2 Interaction-control repeats

1) *Repeats in IRF exchanges*

Though this analysis adopts a CA approach that relies more on an emic perspective that does not abide by any predefined structure, the pervading teacher-centred interaction in most of the examined classes recalls the IRF exchange structure (Sinclair & Coulthard, 1975). In most of the classes and especially in the first fifteen to twenty minutes, the teacher usually presents new language and/ or content items by initiating (I) interaction through elicitation or giving directives. Then, as soon as students respond (R) the teacher gives them feedback (F) and moves to another IRF cycle.

By following this structure, teachers control interaction at this stage of the lesson and keep their students oriented to the pedagogical focus of the lesson. They usually perform repeats of the (I) turn when they do not get the expected response (R) to guide their students into producing the target language that matches the pedagogical focus. Once students perform the right analysis of the pedagogical focus and produce appropriate turns, the teacher provides feedback (F) to close the IRF cycle and starts another one. Excerpt 1 illustrates this kind of repeat with three IRF cycles resulting in a number of repeats of lexical items:

Excerpt 1

- 29 T25: what did you see in the video (1.0) the song
30 T25: yes
31 L: foods
32 T25: *fo:od*, so our *unit* ↑about *fo:od*
33 T25: what's the *title* of our *unit* (1.0) eh
34 T25: what's the *title* of our *unit*
35 T25: yes Abdallah
36 L: the food I eat
37 T25: the *food* I *eat* very *good*
38 T25: the *food* I *eat*
39 T25: what does it *mean* in *Arabic* (1.0) eh
40 T25: what does it *mean* in *Arabic*?
41 T25: yes
42 L: ()
43 T25: very *go:od* excellent

In this excerpt, we can see three teaching exchanges. Each exchange starts with an opening move by the teacher (lines 29, 33 & 34 repeated, 39 & 40 repeated), and ends with a follow-up move (32, 37 & 38 repeated, 43). The repetitions of some of the initiation moves and some of the feedback moves reinforce the teacher's control over interaction, but at the same time multiply the number of tokens and therefore reduce the lexical diversity in those exchanges.

2) Repeats in group work

Controlling interaction is not restricted to teacher-centred phases of the lesson, but sometimes teachers demonstrate control during learners' group work, as well. Teachers try to monitor learners' interaction according to the pedagogical focus as in excerpt 2:

Excerpt 2

- 130 L: teacher finished
- 131 T01: finished?
- 132 T01: Tahani now your turn you choose a *word* [and act]
- 133 L: [I acted (unclear word)]
- 134 T01: OK, do it again do it again
- 135 T01: yes (T. attending to another student)
- 136 T01: OK
- 137 T01: you can give, you can do four *words* for each
- 138 T01: four *words*, at [leats] at least four *words*
- 139 T01: close your books you didn't you don't need a book
- 140 T01: remember the *words* from your memory, please

In this excerpt, the teacher (T01) controls interaction in the group work phase by asking students to use specific words and by setting a minimum number of words that each student should use. In this the teacher performs a number of repeats of the same lexical item *word* to keep her students oriented to the pedagogical focus of the activity which is to practice using some vocabulary items. In fact, by referring to the word frequency list and the key word list of this particular student teacher we find that the lexical item *word* is on the top of both lists.

3) Repeats in repair phases

In some repair activities the teacher performs some repeats to focus a student's attention on a specific language item or structure that needs to be corrected. In that way, the teacher prevents interaction from going in any other direction before the problem is fixed by the student or her classmates. This will be discussed in repair repeats (4.4.4). In addition, control of interaction in repair activities happens when the teacher insists that a selected student corrects the error and prevents other students from helping their classmate. In trying to control interaction in that way, the teacher usually performs repeats as illustrated in excerpt 3:

Excerpt 3

- 333 T02: I am going to study history yesterday
334 T02: yesterday (unclear words)
335 LL: Miss Miss Miss
336 T02: *give* her a *chance*
337 T02: no students *give* her a *chance* (unclear words)
338 LL: Miss Miss Miss
339 T02: you don't know ?
340 T02: OK who will help her ?

By insisting on giving a chance to their classmate, the teacher uses repeats of lexical items *give* and *chance*. Had the students been too insistent on taking turn, it is perfectly plausible that the teacher could have utilised this phase still further, increasing the number of repeats and thus reducing the lexical diversity even more.

4.4.3 Question repeats

1) Elicitation and modification repeats

In elicitation phases of the lesson, teachers use question repeats in order to give their students a chance to think of the answers. At the same time teachers try to modify their questions to help their students find what they are looking for. However, modifications are generally applied to the form of the questions but not to the key words. In other words, key content words are usually repeated in modified questions but function words change. This is illustrated in excerpt 4 below:

Excerpt 4

- 15 T17: OK from this picture what can you *see*
16 T17: what *animals* can you *see*
17 T17 what *animals* can you *see* (1.0)
18 T17: can you tell me what *animals* can you *see*?

In this excerpt, the teacher (T17) employs modification in two ways to her elicitation question in line 15. She realizes that her first question missed a key word which is *animal*, so she inserts

it in two modified question forms that follow (lines 16 and 18) to make it specific and to make sure that she stated her pedagogical focus for that activity in a proper way. However, despite the three different forms the two content key words *see* and *animal* are repeated.

In another excerpt (excerpt 5) that is taken from a class that teaches the students about the three main parts of a writing essay: an introduction, a body and a conclusion, the teacher (T03) tries to elicit what students know about the introduction:

Excerpt 5

131 T03: what type of information that we could include in the *introduction*

132 T03: what do we usually write in the *introduction*

Here again, despite her ability to apply a total modification to the structure of the question and replace the individual lexemes with synonyms, the teacher (T03) does not change the word *introduction*, and instead just reformulates the question. This is due to the fact that *introduction* is a key item of vocabulary to the particular activity and so it is repeated. Changing a key word like this can be counter-productive, in that the new word used to avoid repetition can cause confusion if the students are not particularly familiar with any other alternative.

2) Strict question-repeats

Lightbown and White (1984) indicate that some teachers persist in getting answers to their questions by persistently repeating or modifying them. In my examination of my subjects' data, I noticed that in some cases persistence engenders a strict type of repeats whereby the teacher refrains from making any modifications or giving clues because she believes that the answer is obvious or easy. Strict repeats of questions results in higher rate of lexical items repeats. The example in excerpt 6 is a good illustration of this type of repeat:

Excerpt 6

42 T04: Muthanna, *did you eat donut yesterday?* (2.2)

43 T04: *did you eat donut yesterday?* (2.7)

44 T04: Muthanna! *did you eat donut yesterday?* (2.2)

45 T04: yes or no

Prior to this excerpt, the teacher (T04) gives her students drills on giving short yes/no answers to personal questions in the past tense by saying ‘yes, I did’ or ‘no, I didn’t’. When she addresses her student (Muthanna) in this excerpt, it is clear from her three strict repeats that she expects him to be able to reproduce one of the patterns as a result of the drill. However, after getting no response despite the three repeats, she starts to reduce her expectation and she asks for the minimum ‘yes’ or ‘no’ in order for her to start scaffolding after that.

3) *Think-time Repeats*

Think-time or wait-time is the time given by the teacher to her students to think of an answer for a given question. It is usually reported that teachers do not give think-time for their students. Walsh (2013, p.122) confirms this reality by saying that ‘typically-it is less than one second’. In the case of the present study, I observe that most of the subjects usually ask a question at least twice before they designate a student to answer even when they already have volunteers before the repeat. My interpretation of the repeat in this case is that teachers use it as a think-time for the students, and though it is not a silent one as it is expected to be, the repeat can help the students to process the question as they listen to it for a second time. Excerpt 7 illustrates how teachers use repeat for think-time:

Excerpt 7

- 50 T05: who *wants* to *play* first
51 T05: who *wants* to *play* first
52 T05: Noura come here
53 T05: listen to the question
54 T05: what’s the *colour*?
55 T05: what’s the *colour*?
56 L: pink
57 T05: yes
58 T05: where is the *pink* Noura
59 T05: where is the *pink*
60 T05: read out the question

Apart from this one-time repeat, teachers sometimes use many repeats to extend think-time. In excerpt 8 below, the teacher (T14) discusses a reading passage with her students where a grandfather tells his grandson about life in the past. In the while-reading stage, the teacher asks her students to guess what the grandfather would say when his grandson asks him about the way people used to travel from one place to another:

Excerpt 8

- 348 T14: what do you *think*
- 349 T14: what do you *think*
- 350 T14: how did *people* use to *travel* from place to another=
- 351 LL: =teacher [teacher, teacher, teacher, teacher, teacher, teacher]
- 352 T14: [in past what do you *think* the *grandfather* will say]
- 353 T14: what do you *think*
- 354 T14: yes WHAT DO YOU *THINK* =
- 355 LL: =teacher [teacher, teacher, teacher, teacher]
- 356 T14: [how did they *travel* in *past* ?]
- 357 T14: [what do you *think* the *grandfather* will say ?]
- 358 LL: [teacher, teacher, teacher, teacher, teacher]
- 359 T14: yes [what do you *think*]
- 360 LL: [teacher, teacher] teacher, teacher, teacher
- 359 T14: Mozah Darwish stand up
- 360 T14: what do you *think* what do you *think* the *grandfather* will say
- 361 T14: how did *people* used to *travel* in *past*

In this excerpt, the teacher had volunteers to answer her question (line 351) sooner than she has first asked it. However, she preferred to extend the think-time for slower students by repeating the question ‘what do you think?’ five times before choosing a student to answer (line 359). I believe that her repeats were meant for the slow ones or the ones who did not raise their hands

to demonstrate their willingness or ability to answer. The way she raised her voice in line 354 indicates that she was urging them to participate. In total, the teacher repeated ‘what do you think?’ nine times in this sequence which increases the effect of the content word ‘*think*’ on reducing the lexical diversity of this subject in class.

On another note, what I noticed with this specific teacher (T14) is that she uses no repeat or just one-repeat with display questions as she did when she reviewed names of inventions with her students (appendix 94). However, when she started the reading and she started asking referential questions that required more thinking, she used extended think-time repeats.

4.4.4 Feedback repeats

1) Confirmation repeats

In confirmation repeats, the teacher repeats students’ answers once or more times to confirm the correctness of their answers and as a form of reinforcement. Park (2014) indicates that a teacher’s repeat in the third turn can be performed to confirm a correct answer or a contribution by a student in turn two. However, this type of repeat can be meaningful for the actual study only if the teacher makes more than one repetition of the student’s answer as the focus is on repeats that multiply the number of tokens which in turn reduce the values of lexical diversity. In my examination of the subjects’ recordings and transcripts, I found that some teachers perform this type of repeat more than one time especially when they collect answers from different students as in excerpt 9 below. In this excerpt, the teacher (T10) asks her students to move around the room and show her things that have the colour that the teacher mentions:

Excerpt 9

- 79 T10: >yellow, yellow, where is yellow <
80 LL: [teacher teacher]
81 T10: [go there] (5.0)
82 LL: teacher, teacher, teacher, teacher
83 T10: this is yellow ()
84 T10: ↑yellow (2.0)
85 T10: yes kids ↑yellow (2.0)

- 86 T10: HERE, HERE (3.0)
87 L: teacher yellow, yellow
88 T10: OK show me, show me *green* =
89 LL: =green
90 T10: >*green*, *green*<
91 LL: teacher, teacher, teacher
92 T10: this is *green*, excellent *green*
93 T10: *green* yes, excellent yes
94 T10: excellent T03ah *green*

In this excerpt the teacher (T10) confirms her students' actions of showing her the right object with the named colour by performing a number of repeats.

2) *Praise-word repeats*

Praising words are a common feature of teacher talk that they use after a student's correct answer or appropriate contribution to keep them motivated. Some teachers vary their lexical items according to the quality of the student's answer or according to the effort made by the student in performing some activity. It is noticed that the subjects in this study usually keep using and repeating the same type of lexical items for praise throughout the class. By looking back at the generic class list of the most frequent words used by the subjects (appendix 9), we can see that *excellent* and *good* are on the top of class word list with *excellent* occurring 365 times and *good* occurring 335 times. This pervasive repetition of lexical items like *excellent* or *good* in most of the classes help student teachers to monitor sequence organization in order to keep interaction oriented in a normative way to the pedagogical focus of the lesson. However, those repetitions cause lexical diversity scores to decrease due to the high number of tokens. The following example (excerpt 10) illustrates the frequency with which one of the subjects uses *excellent* in class interaction. The teacher is checking her students' knowledge of subject-verb agreement before using 'going to' to talk about the future:

Excerpt 10

- 96 T02: if I want to say they
97 LL: Miss, Miss, Miss, Miss, Miss
98 T02: yes
99 L: they are
100 T02: they ↑are, *excellent!* because they are ↑they ARE
101 T02: many .hh there's not - not one ()
102 LL: plural
103 T02: yes, *excellent!* plural, we, we
104 LL: Miss, Miss, Miss, Miss, Miss
105 T02: we are, also because we ↑is
106 LL: plural
107 T02: pl - plural also, *excellent!*
108 T02: what else
109 T02: and I?
110 LL: Miss, Miss, Miss, Miss, Miss
111 L: am, I am
112 T02: I am, *excellent!* ↑I AM
113 T02: so you know student
114 T02: ve:ry good, *excellent!*

Notice that between lines 100 and 114, *excellent* was repeated 5 times (lines 100, 103, 107, 112 & 114). When we take a look at the frequency word list of this specific subject we find that *excellent* is the highest content word -after the semi-content word *be-* that she uses in 52 occurrences (appendix 11). In the case of another subject (T16), it is noticeable that she uses both *excellent* and '*clap your hands for* (a name/ your friend)' in an abundant way that her word frequency list shows that *excellent* is number is the second most frequent word after *be* with 46 repeats, *clap* is ranked the fourth with 23 repeats, and *hand* is number ranked fifth with 22 counts (appendix 25).

Though student teachers can use other lexical items such as *perfect, fantastic, good job, well done*, among others to praise their students, they keep using one or two types of lexical items because they focus on the pedagogical focus of their classes rather than on displaying their lexical diversity.

3) *Repair repeats*

Like questions, repair occupies a considerable amount of a teacher's talk. Van Lier (1988) considers repair as another important characteristic of classroom language apart from questions. Some of the teachers opt for other-initiated self-repair strategy to give students a chance to correct their own mistakes. They sometimes repeat students' wrong answers more than once with a rising intonation to draw their students' attention to the mistake in order for them to correct it by themselves, and to avoid negative verbal feedback (Seedhouse, 2004). Excerpt 16 illustrates this kind of repeat. In this lesson the teacher (T08) discusses the theme 'movies' with her students and asks them why people eat popcorn when watching movies:

Excerpt 11

- 380 L: when we be angry
381 T08: when we be ↑ *angry*
382 T08 ha::: ((T produces an irate sound/ acts as angry person))
383 T08: we are *angry*?
384 L: hungry
385 T08: *hungry*?
386 T08 ↓OK::
387 T08: so when you are *hungry* you eat popcorn ()

In this excerpt, the teacher initiates repair through a repetition of the student's wrong answer in two stages. In line 381, she repeats her student's answer verbatim, but she raises her tone at the word *angry*, and she couples her repetition with non-verbal clues (irate sound and facial expressions) in line 382. At the same time, the teacher ignores the wrong form of the verb *to be* in her first repeat to keep her student's attention focused on the content word *angry*. However, in her second repeat in line 384 and before her student starts self-repair she corrects the form of

to be in preparation for her student to produce a fully-repaired sentence. As indicated by Seedhouse (2004, p.149), the teacher's pedagogical focus which is on the meaning and not on the form made the teacher in this case give priority to self-correction of the content word, and she avoided confusing her student with correction of the form as well. Instead, she conducted embedded correction.

In some other cases teachers opt for other-initiated other-repair. In both cases, teachers perform repetition of content vocabulary that reduce the values of their lexical diversity, especially when error correction happens with low-achieving students who require more repetitions to manage to apply the correction.

4.4.5 Key-word repeats

1) Lesson key-word repeats

This type of repeat is common to all of the examined lessons. As teachers and students interact, they keep using lexical items that relate directly to the main focus of the lesson. They are key word(s) in the usual sense rather than the technical corpus linguistics sense that represent the common thread that preserves the coherence of interactional sequences. The following excerpts are taken from a lesson that focuses on the difference between adjectives and nouns and the position of adjectives in a simple sentence. In excerpt 12, which happens at the beginning of the class, the teacher clarifies the difference between an adjective and a noun and gets examples from her students.

Excerpt 12

41 T18: OK now (1.3) I want you to tell me what's the differences between **noun**

42 and ad .hh and **adjective**

43 T18: what's the differences between ↑**noun** and **adjective**

44 L: uh?

45 T18: what's the difference <between **no:un** (3.0) and **adjective**>

46 L: teacher (9.0)

47 T18: yes Asma (2.0)

- 48 T18: [yes]
- 49 L: [اسم] ((tr.: noun))
- 50 T18: yes the *noun* means الاسم ((tr.: noun)) (.)
- 51 T18: *adjective* means صفة ((Tr.: adjective)) (.) OK
- 52 T18: give me any *noun* related to the theme
- 53 T18: what's our theme about
- 54 L: nature
- 52 T18: nature
- 53 T18: give me any *noun* related to nature
- 54 T18: yes Shamsa
- 55 L: sun
- 56 T18: sun excellent (1.0)
- 57 T18: what else
- 58 L: yellow in eh adj .hh
- 59 T18: for *adjective* yellow yes excellent sit down Shamsa

In this excerpt of 18 lines, the teacher repeats the word *noun* six times and *adjective* five times as she explains their meaning and as she elicits examples from her students. Then, in the next phase of the lesson, the teacher gives her students some practice activities. The following excerpt illustrates how the same key words are repeated a number of times by the teacher as she explains the activity to her students:

Excerpt 13

- 106 T18: OK listen you have these word in the box you have to (.) categorize them
- 107 to *adjective* and *noun* (.) OK?
- 108 T18: put the *noun* in the box of *noun* and put the *adjective* in the *adjective*
- 109 T18 for example you will put (guess) black in the *adjective* ()
- 110 T18: and cloud in the *noun* (.) OK?
- 111 T18: you have two minutes to do it (.) you have two minutes

In this short sequence, the teacher repeats noun four times and adjective four times in just four lines (lines 107-110) as she tries to give clear instruction for her students to carry out the assigned task in a successful way.

Later on, when she corrects the task with her students, she demonstrates to them (on the board) how they should write the nouns and adjectives under their corresponding columns. In the following excerpt (excerpt 14), she tries to clarify that they are starting with the adjective column:

Excerpt 14

144 T18: OK I put the ↑black (3.0) in this column

145 T18: I put in the *adjective* column (.) LOOK (2.0)

146 T18: this one is for *adjective* LOOK (1.0)

147 T18: this one for *adjective* and this one for *noun*

148 T18: now we will use the *adjective* (.) we will try to do the *adjective*

In this sequence, the teacher repeats *adjective* five times to make sure that they are focused on the adjective column.

In fact, by looking at the list of key words generated by Wordsmith for this subject (T18), we can notice that *noun* is ranked number one and *adjective* is ranked number two (appendix 71) which obviously corresponds with the evidence provided by the above-mentioned excerpts that are taken from three different phases of the lesson. Similarly, the word frequency list shows that *noun* is the second most frequent content word after the word *sentence*, and *adjective* is the fourth after the word *number* (appendix 27). While these teacher repeats fulfil a desirable pedagogical purpose of keeping interaction oriented to the focus of the lesson, they play a restricting role on the lexical choices of the subjects. Therefore, teacher talk remains within a specific lexical range that does not require the display of lexical diversity but shows a pedagogical skill that is conducive to effective learning.

2) Activity key-word repeats

Activity key words are repeated by the teacher for the time a specific activity is conducted to keep the students oriented to the pedagogical focus of that activity. They are not like the lesson key words because they appear only when a specific activity starts and they usually disappear as it finishes. In other words, the teacher starts using them as he/she states the pedagogical focus of an activity, and as he/she gives instructions, then stops using them as soon as the activity finishes. In excerpt 15, the teacher (T23) plays a guessing game with her students that she calls 'Riddles' to practice using vocabulary related to materials like metal, paper, wood, glass, and so on. She invites individual students to pick folded papers and read the riddle to their classmates then make a guess:

Excerpt 15

38 T23: OK girls first we want to play a game OK?

39 T23: do you like *riddles*?

40 LL: yes

41 T23: do you like the *riddles*?

42 LL: ↑yes

43 T23: what is the *riddle*

44 T23: what is the *riddle*

45 T23: what is the meaning of the *riddle*

46 LL: teacher teacher teacher

(Omitted six lines where the teacher was dealing with a classroom management issue)

53 T23: what is the meaning of *riddle*?

54 L1: ()

55 T23: excell:ent thank you Shahad (2.0)

56 T23: OK now I want someone to pick one paper (.) from the box (.) and read

57 the *riddle* (1.4)

58 T23: after that (.) she has to answer (.) the *riddle* OK?

As noticed, to keep her students focused on the aim of the activity which is to solve riddles the teacher uses eight repeats of the word *riddle*. However, as the activity starts the word *riddle* disappears and lexical items that convey the meaning of guessing appear in teacher repeats. The written riddles follow the same pattern: ‘I am made of (material), you use me to (do something)’. While the selected students think of the answer, the teacher keeps repeating the riddle as illustrated in excerpt 16:

Excerpt 16

- 85 T23: did you understand? (1.0)
86 T23: I am *made* of *wood* (2.3)
87 T23: I am *made* of *wood*
88 T23: you *use* me to *put things* in (2.0)
89 T23: what is this
90 T23: I am ↑a
91 T23: yes Nouf
92 L: () (3.0)
93 T23: it is something (.) you *use* it to *put things* in
94 T23: it is *ma:de* of *wood*

As illustrated in excerpt 5, the teacher keeps repeating the highlighted content words and this recurs with all the students who take part in the game. By looking at the key words list generated by Wordsmith for this subject (T23), we notice that the word *made* is the second on the list after the word *excellent*. This can be explained by the fact that the activity continued for about nine minutes and the teacher kept repeating especially the second line of the riddle ‘I am *made* of ...’.

Once the ‘Riddle game’ activity is finished that kind of repetition stops and a new one starts with a new activity. The activity that follows is solving a crossword puzzle that focuses on inventions. Therefore, the key word for the activity is *invention*. As in the previous activity, the teacher (T23) keeps repeating the word *invention* to keep her students oriented to the

pedagogical focus of that activity. However, to avoid overloading this section with similar excerpts, I refer the reader to appendix (101), lines 228-245 to see teacher repeats of the word *invention*. The key word list for this subject shows that *invention* is the third key word after *excellent* and *made* (appendix 76).

3) *Story key-word repeats*

In story telling classes for young learners, repeats of rhyming lines and of key words are very common. Teachers perform those repeats as part of their reading of the stories for their students, but they also engage their students through multiple repeats and they invite their students to repeat after them. In excerpt 17, the teachers reads ‘We’re Going on a Bear Hunt’ story for her students and performs a number of repeats:

Excerpt 17

- 107 T20: I am going to read first and then you will read with me OK
- 108 T20: we’re **going** on a **be:ar** **hunt** =
- 109 LL: =we’re going on a bear hunt=
- 110 T20: =we’re **going** on a **bear** **HUNT**=
- 111 LL: =we’re going on a bear hunt=
- 112 T20: =we’re **going** to **catch** a **big** **ONE** =
- 113 LL: =we’re going to catch a big one=
- 114 T20: =we’re **going** to **catch** a **BIG** **one** =
- 115 LL: =we’re going to catch a BIG one

4.4.6 Approach-related repeats

1) *Language-drill repeats*

Language drill repeats are common in L2 classrooms despite the differences in frequencies with which they practised. This variation emanates from the teaching and learning methods that teachers use. They are usually repeats of new words or new grammar forms and patterns that

teachers introduce in classes. Vocabulary drills usually focus on improving students' pronunciation of new words or helping memorization. Excerpt 18 is an example of vocabulary drills:

Excerpt 18

183 T16: we learn the numbers from one to ten and today we will continue

184 from eleven to twenty, OK

185 T16: so who can read this ?

186 LL: eleven

187 T16: *eleven*

188 LL: eleven

189 T16: all of you *eleven*

190 LL: eleven

191 T16: *eleven*

192 LL: eleven

193 T16: not elevant

194 T16: *eleven*

195 LL: eleven

196 T16: *eleven*

197 LL: eleven

198 T16: all of you

199 LL: eleven

200 T16: I cannot hear you

201 LL: eleven

202 LL: eleven

203 T16: *eleven*

- 204 LL: eleven
205 T16: *eleven*
206 LL: eleven
207 T16: excellent

In this excerpt the teacher (T16) introduces a new number and keeps repeating it to model the right pronunciation. That's why in line 193 she corrects the pronunciation of some of her students (eleven not 'elean').

In some other cases drills are used to help students memorize some grammar forms or patterns as illustrated in the following excerpt

Excerpt 19:

- 125 T06: long we do not say bigger we do not say tall to hair we say ?
126 LL: long
127 T06: so Fatmah's hair is what ?
128 LL: longer
129 T06: everybody *longer*
130 LL: longer
131 T06: *longer*
132 LL: longer
133 T06: very good

2) Scaffolding repeats

Peregoy and Boyle (2005, p.100) indicate that scaffolding is built on support, encouragement and assistance provided by teachers for their learners as they challenge them so they move to the next level development. Scaffolding is another important feature of L2 classroom. Teachers join their students' efforts in communicating their messages by providing support while maintaining a reasonable level of challenge that keeps students engaged. Walsh (2011) mentions three types of scaffolding, namely, *latched modelling*, *alternative phrasing*, and *prompting*. In

latched modelling the teacher tries to model the target language form following incorrect or incomplete turns by students to guide them into producing it correctly. This takes some repeats until the students manage to produce the target form. The teacher (T12) in excerpt 20 performs latched modelling as she tries to get her student to produce the form ‘I like to eat ...’

Excerpt 20

- 391 T25: what do you *like* to *eat* for the breakfast
392 T25: yes Saif
393 L: cornflakes and egg
394 T25: full sentence
395 L: I eat=
395 T25: =I *like*
396 L: =I like cornflakes and egg
397 T25: I *like* to *e:at*
398 L: °I like to eat°
399: T25: cornflakes
400 L: and egg=
401 T25: =and an egg very good, OK

In this excerpt the teacher performs latched scaffolding with minimal repeats and she manages to get the student to produce the target pattern ‘I like to eat’. However, in excerpt 18 below the teacher (T12) fails to get the student to produce the target pattern ‘I want’ though she combines latching and prompts which results in lots of repeats of the same lexical item ‘want’:

Excerpt 21

- 145 T12: yes Ali
146 T12: I
147 L: I is
148 T12: *want*
149 T12: choose something from here ((teacher showing pictures of food items))

- 150 T12: what do you **want** ?
- 151 T12: boys listen to Ali
- 152 L: xxx .
- 153 T12: yes Ali
- 154 T12: I **want**
- 155 T12: what do you **want** ?
- 156 T12: do you **want** ice-cream ?
- 157 T12: do you **want** cake ?
- 158 T12: do you **want** beef-burger ?
- 159 T12: do you **want** pizza ?
- 160 T12: do you **want** milk ?
- 161 T12: do you **want** hotdog ?
- 162 T12: what do you **want** Ali ?

(Lines omitted where the teacher asks another student to model the right form)

- 172 T12: Ali what do you **want** ?
- 173 T12: you **want** ice-cream ?
- 174 T12: yes excellent clap your hand for your friend

As noticed in the excerpt above (excerpt 18), the teacher (T12), uses three different ways in scaffolding in an attempt to help Ali (her student) to produce the pattern 'I want + food item'. She uses repeated prompts (lines 156-161). She uses visual clues (line 149). She even uses one of Ali's classmates to model the right form. Finally as she notices that Ali faces clear difficulty in producing the pattern, she decides to close this scaffolding sequence by accepting a non-verbal response from Ali showing that he wants ice-cream (lines 173 & 174).

4.4.7 Procedural repeats

1) Instruction-clarification repeats

Procedural repeats happen especially when the teacher explains the procedure for carrying out an activity. Teachers usually tend to repeat their statements to make sure that students

understand instructions, and they usually repeat key words more than any other words to draw students' attention to them. Excerpt 22 illustrates this kind of repeat:

Excerpt 22

- 44 T15: do you want to play a game?
- 45 LL: yes
- 46 T15: OK now <listen to me> listen to me OK?
- 47 T15: listen I will **choose** ten **students** (.) ten **students**
- 48 T15: five **student** from this **group** (.) and five **student** <from this **group**>
- 49 T15: and I will **choose** one **judge**
- 50 T15: one **judge** OK?
- 51 T15: then (1.0) the five **students** **come** here and **make** <one **line** one **line**> OK?
- 52 T15: the first for this **students** **come** here and **make** (.) one **line** OK?
- 53 T15: then I have a **paper** with **sentences** (3.2)
- 54 T15: OK (1.0) a **paper** with **sentences** (.)
- 55 T15: the **judge** will **read** the **sentences** OK?
- 56 T15: and then **one** **student** from **group** one will **come**
- 57 T15: and **one** **student** from **group** two will **come** here
- 58 T15: then the **judge** will **say** ↑the **sentence** (.) will **say** the **sentence**
- 59 T15: the **student** **act** **act** the **sentence** OK?
- 60 T15: the **student** who who **act** fast (.) her team will get one point OK?
- 61 LL: OK

In this excerpt the teacher (T15) repeats eleven content words to explain the procedure for the activity. She performs different counts of repeats for each content word depending on their importance for the activity. The word *student* gets the highest count of ten repeats which indicates the teacher's intent of making the activity student-centred. It also indicates that the teacher wants her students to realize that they have an active role in making it a successful activity. It is clear, then, that *student* is a key word in this procedural sequence. If we take a look

at the key-word list of this subject (T15), we notice that, in fact, the word *student* is on the top of the list (appendix 68). However, if we look at the generic class word-frequency list (appendix 9), we find that it is not on that list, which confirms that those repeats reflect the centrality of the word *student* for the procedure of that specific activity in order for the teacher to explain it clearly .

To make sure that this type of repeat is not restricted to complex procedures as in excerpt 22, I examined other examples which are simpler, and I found that procedural repeats are frequent in both cases. Excerpt 23 is an example of simple procedural sequence that uses repeats of the same lexical items:

Excerpt 23

- 175 T22: OK it is a game OK?
176 T22: and you are four **groups**
177 T22: I will give each **group** a question
178 T22: and you have to **answer**
179 T22: if your **answer** is correct you ↑will **play**
180 T22: you will continue **playing**
181 T22: and if not? you will stop (.) you will not **play** again OK?
182 T22: in **groups** you have to discuss the answer before you give it to me OK?

Despite the simple procedure that is explained by the teacher in this sequence, we notice repeats of three lexical items that are keys for students' understanding of this simple procedure. This confirms that repeats of lexical items in procedural sequences are a necessity for the teachers to make sure their students understand what is required from them .

Another form of procedural repeat that occurs in a number of the examined classes is for clarification of instruction when the teacher notices that her students or some of them are doing the assigned activity in a wrong way. In excerpt 24, the teacher (T22) realizes that her students are writing sentences instead of drawing as she has instructed them to do.

Excerpt 24

- 476 T22: there is one orange on the table
- 477 T22: you have to *draw* one orange on the table
- 478 T22: no just *draw*
- 479 T22: *don't write* the sentence ↑girls ((teacher claps her hands))
- 480 T22: *don't write* the sentence just *draw*
- 481 L1: نرسم ((Tr.: we draw))
- 482 T22: yes *draw*
- 483 L2: نرسم ? ((tr.: draw?))
- 484 T22: ye::s *draw*

The teacher (T22) makes five repeats of the word *draw* as a key word for this activity to draw her students' attention to the wrong practice that she has observed with some of them. She use two repeats of '*don't write*' to make it clear that there is no writing involved in the activity, and she carries on with the repeat of '*just draw*' for the whole time of the activity. In fact, The key word list of this subject shows that *draw* is the second on the list. This kind of repeat is also necessary in a clarification sequence so students can do the activity properly. It is noticed that even with the repeats of *draw* and *just draw*, two students needed confirmation that they understood it correctly by using their mother tongue.

2) Classroom management repeats

Classroom management is an essential feature of a successful lesson. Scrivener (2011, p.54) claims that 'the skills of creating and managing a successful class may be the key to the whole success of a course'. Teachers tend to use different classroom management techniques to make sure their classes run smoothly to achieve the planned objectives. What is noticeable about the examined transcripts is the number of repeats that most of the teachers perform in stating or repeating classroom rules and in giving instructions that aim at regulating students' behaviour. Some of those repeats are performed at the beginning of the activity to make sure that students start at the same time. Excerpt 25 illustrates this kind of repeat:

Excerpt 25

- 38 T07: OK now I will give (3.0) each one a card (20.0)

- 39 T07: *wait* I will give you (7.0)
- 40 T07: *wait, wait* don't *start*
- 41 (unintelligible student-student interaction) (8.0)
- 42 T07: ↑*wait, wait* don't *start*
- 43 (unintelligible student-student interaction) (7.0)
- 44 T07: OK now *start* (.) boys *start*

In this excerpt, the teacher (T07) uses repeats of two content words: *wait* and *start* to manage the beginning of the activity in a way that all students start at the same time in order for the activity to achieve its aim.

Another type of classroom management repeat is performed to manage students' movement and behaviour especially during class and group activities. In excerpt 26, the teacher (T08) tries to manage the behaviour of some of her students during a silent reading activity:

Excerpt 26

- 582 T08: don't *tell* me teacher
- 583 T08: *sit down, sit down, sit down, sit down*, don't talk
- 584 T08: others are reading OK
- 585 T08: >*respect* yourself, *respect* your friends< and one of you stand up and *tell*
- 586 me in English

In this excerpt, the teacher (T08) gives different instructions to manage some of her students behaviour and uses repeats to emphasize the behaviour that those students should display during a silent reading time. Her four successive repeats of the content word *sit* indicate her concern with the fact that some students were distracting their classmates who were reading by standing up and talking to her in Arabic.

The third kind of repeats that is usually happening in classrooms is when the teacher tries to get students' attention to her. Excerpt 27 is an example of this kind:

Excerpt 27

- 563 T26: all the class

- 564 T26: *listen* ↑cla:ss
- 565 T26: group (Teacher claps) ↑one yellow ↑team ↑yellow
- 566 T26: wait, wait Fatmah yellow and pink
- 567 T26: *listen* to me *listen* to me (2.0)
- 568 T26: CLASS *listen* to ME
- 569 T26: blue TE:AM here
- 570 T26: *listen* I will read and you try to answer with your friend OK
- 571 T26: *listen*

In excerpt 27, the teacher produces six repeats of the word *listen* along with a loud voice (marked in capital letters) to try to attract her students' attention to what she was going to read for them. By looking at the list of the most frequent words of this subject, we can notice that *listen* figures in the top ten words, which indicates that it is a key word in classroom management for this teacher. In reality, the generic word frequency list (appendix 9) shows also that *listen* figures in the top ten frequent words for all the subjects. This tells that it is one of the most influential lexical items that reduce the subjects' diversity in the classroom.

The last sub-type of classroom management repeats is the one used to mark the end of an activity. Teachers usually perform many repeats of the word *finish* to remind students of the finishing time or to check that they have finished. Excerpt 28 illustrates this kind of repeat:

Excerpt 28

- 524 T10: *finished* Fatmah? (5.0)
- 525 L: teacher
- 526 T10: Maryam sit down (10.0)
- 527 L: teacher finish=
- 528 T10: =Wisal sit do:wn (9.0)
- 529 T10: *finished* here? (6.0)
- 530 T10: Maithah quickly *finish*
- 531 L: finish teacher

532 T10: have you *finished* ?

533 LL: yes .

534 T10: *finished* Maithah ?

In this excerpt, the class comes to an end and the teacher (T10) uses five repeats of the word *finish* to make sure that her students finish their task before leaving the room, but also as a way to urge those who have not finished to complete their work before the end of the class.

4.5 Conclusion

In this chapter, the collected data was analysed quantitatively and qualitatively. The quantitative analysis was based on data driven from three different tools. First, VOC-D was used to compute the lexical diversity values for each subject. Lexical diversity was used as a viable indicator of language proficiency as it has been proved by the reviewed literature. Second, SPSS computed the means, the standard deviation, and three correlations: (1) between the two sets of D-scores for the IELTS speaking test and the class, (2) between the MCT grades and IELTS scores, and (3) between the MCT grades and class D-scores. The comparison between the D-scores, the means and the standard deviations showed that there were clear discrepancies between the two performances of the subjects on the IELTS speaking test and in the class. Causes were attributed mainly to differences in contexts and related language requirements. The three correlations were identical in that they revealed no relationship between the different variables. The conclusion that is drawn from the findings of the first two correlations is that the subjects' performance on the IELTS test does not correspond with their performance in the classroom. The assumption is that the test elicits different vocabulary and different language than the classroom because there is more for the teacher to focus on in the classroom than just the language. The third correlation which was, also, proven to be insignificant consolidated the idea that the language factor is only a small component of the criteria for which student teachers' performance is evaluated. Moreover, Wordsmith tool generated word frequency lists for each subject and for the whole group. This quantitative data permitted to see clear discrepancies between the subjects' lexical diversity on the IELTS speaking test and in class at the level of the most frequent content-words used in each context. Word lists showed reliance mainly on verbs in class and more on nouns

on the speaking test. This finding geared the qualitative analysis towards seeking factors that make classroom talk less diverse than the test. A close examination of the transcripts using corpus linguistics and Conversation Analysis revealed the important impact of word repetitions performed by the student teachers upon the values of their lexical diversity. A taxonomy of 6 categories and 16 types of repeats that teachers use in the classroom has been created and analysed with reference to examples from the subjects' classes. The conclusion drawn is that a teacher's language in the classroom is very much controlled by different features of classroom interaction, among which word repetitions have been identified as the strongest in this research study.

Chapter 5: Discussion of the findings

5.1 Introduction

The aim of this chapter is to discuss the findings of the present research study in light of the literature review presented in chapter 2, and with reference to the methodology outlined in chapter 3, and the data analysis discussed in chapter 4. It is hoped that the forthcoming discussion will contribute to the body of knowledge available on language proficiency testing and its relation with classroom interaction. In this chapter, I will present a triangulated discussion drawing on the complex relationship between the different components of three main concepts, namely language proficiency, classroom discourse and language testing. The discussion will follow an integrative approach that interprets the quantitative and qualitative findings with reference to the different models of language proficiency, the features of classroom interaction, and the types of validity. While the purpose is to answer the main question of the present research study regarding the appropriateness of using IELTS scores to predict student teachers' performance in the classroom, each section will discuss the findings presented in chapter 4 in relation to one of the sub-questions suggested in chapter 1.

The main question of this research study is:

Can IELTS test scores make accurate predictions of student teachers' performance in the classroom?

The sub-questions are:

- 1) Do student teachers display similar lexical diversity on the IELTS speaking test and in the classroom?
- 2) Do IELTS scores match college mentors' grades for student teachers on teaching practicum?
- 3) How does classroom interaction affect student teachers' lexical diversity scores?

5.2 Do student teachers display similar lexical diversity on the IELTS speaking test and in the classroom?

Based on the statistical analysis presented in the previous chapter, figures show that there are clear discrepancies between students' lexical diversity scores on the IELTS speaking test and in the classroom. A simple comparison between the two sets of lexical diversity D-scores reveals that more than 80% of the subjects displayed higher levels of lexical diversity on the test than in the classroom. The comparison between the overall means gives clear evidence that performance on the IELTS test is significantly higher than performance in the classroom with a mean difference exceeding 12 points. Moreover, a comparison of the standard deviation on IELTS and classroom reveals a much bigger spread of scores in classroom lexical diversity than on the test. This can be attributed to the differences in interaction and the types of addressees, which will be discussed later in this section. Finally, Pearson correlation test confirms the initial hypothesis that there is no relationship between the D-scores on the test and those in the classroom as the correlation coefficient is (-.160), and significance is (.425). So, how can these discrepancies be explained?

With reference to Bachman and Palmer's (1996) framework of language proficiency and language testing, it would be appropriate to discuss the relationship between the language test task which is the IELTS speaking test, and the language use task which is classroom teaching in order to gauge the degree of correspondence and its impact on the lexical diversity displayed by the subjects in this study.

5.2.1 The correspondence between language teaching and the IELTS speaking test

In their conceptual framework, Bachman and Palmer (1996) emphasize the correspondence between the characteristics of the language use task and the language test task. In the case of the present study, the language use task is teaching in the classroom, whereas the language test task is performance on the IELTS speaking test. The lack of correlation between the lexical diversity scores of the same subjects revealed by the statistical analysis in the previous chapter (4.2), suggests that there is a lack of correspondence between the classroom teaching task and the IELTS speaking test task. This implies that the two tasks have different characteristics that

compelled the subjects to display lexical diversity that was in some cases twice as high on the test as in the classroom (T04 & T17 in table 4.1).

One of the major characteristics of the task of teaching young L2 learners in the UAE schools where English is not used as the means of instruction in other subjects is the use of basic vocabulary that is within the Zone of Proximal Development of those young learners. Therefore, it is expected that the subjects in this study use a limited range of high frequency words that their students are familiar with, and they are expected to introduce new words only as long as they are part of the objectives of their lessons. Moreover, student teachers simplify their speech by using the most common vocabulary, and they repeat words in the classroom for different reasons -that will be discussed later on- in order to make sure that learning is taking place. By contrast, on the IELTS speaking test, candidates try to adjust their language level to their audience, in this case, the examiner. In the test context, they are aware that their addressee has a high level of vocabulary knowledge, so they try to vary their lexical items and avoid repetitions to try to give the examiner a good impression about their lexical diversity. They also find themselves compelled to use more than just the basic vocabulary that they use in the classroom to be able to cope with the linguistic demands of the interview questions. This happens especially in the third part of the IELTS speaking test which requires higher-order thinking (see 2.5.4), hence they use advanced vocabulary. This is, indeed, reflected in the statistical findings revealed by Wordsmith tools (see 4.3.1) which indicate that there is a clear difference in the type of lexical diversity resulting from each task. As a matter of fact, 70% of the subjects' top twenty frequent words in the classroom task, are verbs (figure 4.1), which can be explained by the fact that their language task in the classroom is different as they act as teachers who initiate interaction and who give frequent instructions to their students to carry out tasks and planned activities. The concordances of the most frequent verbs like *write*, *go*, *come*, *sit*, *read*, *listen*, and *look* (appendices 37-47) show that they are mostly used to give instructions (see verb lemmas in 4.3.2). Moreover, the classroom language task is marked by the use of the adjectives *excellent* and *good*. These appear at the top of the class content-word frequency list (appendix 9). They indicate a fundamental characteristic of the classroom task which relies heavily on praising learners and encouraging them continuously. Conversely, *excellent* does not appear on the IELTS most frequent words, and *good* appears at the bottom of the list not as a praise word but as an adjective for the description of activities, events, and so on.

The IELTS word list shows that there is more use of nouns and less of verbs with a room for adverbs, as well, that do not appear on the class list. This makes the IELTS list show more diversity compared to the class list. The two tasks in this sense do not seem to be going in the same directions. This might arguably account for the negative coefficient (-.160) as the L2 classroom task requires less diversity and more frequent words to facilitate communication and learning, and the test task requires more diversity and less frequent words for two main reasons: (1) to meet the requirements of a high score, and (2) to cope with the difficulty level of the questions especially in the third part of the interview which requires higher-order thinking and knowledge of some abstract vocabulary. This leads us to discuss the cognitive demands imposed by each task on the subjects while performing them.

1) *Task-related cognitive demands and context embeddedness*

Following Cummins's (2000) framework of language proficiency, there seem to be clear discrepancies between the classroom teaching task and the IELTS speaking task. In the classroom, the cognitive demands on the subjects are not at a high level as far as language is concerned. It is an embedded context as the subjects are the 'owners' of the class. They plan and they prepare for the teaching task in advance. They are the ones who decide on the content and the level of language use in the classroom according to their academic goals, the pedagogical focus, and the level of their students. Their preparation for the different activities that they plan to conduct in the classroom makes them aware of the lexical range they can use to achieve the learning objectives. It is this factor that makes the task undemanding as they are not required to use their vocabulary knowledge at its optimal level while carrying out the task, but rather to limit their language use to the vocabulary that best fits the level of their students and the goals of the lesson.

By referring to Cummins's framework (figure 2.2), the language use task in the classroom would fit into quadrant A where the context is embedded and the cognitive demands of the task are not really at a high level. However, the IELTS speaking task is context-disembedded as the test questions provide no contextual support for the language test takers, which puts more pressure on their cognitive abilities and renders the task cognitively demanding. Unlike the language task in the classroom, the subjects can do no specific preparation for the test task because the questions are unknown until they sit for the test. The first part of the test remains the least

demanding because it is about familiar topics for the candidates, like talking about themselves, their homes, their families, their jobs or studies and so on. In part two, the task is cognitively demanding and the context is disembedded as the candidate is given a randomly-chosen topic and only one minute to plan for a one-to-two-minute talk. The lack of contextual clues and the time limit for the preparation put candidates under pressure as they need to develop an appropriate answer while using the best lexical range to articulate their ideas within the planned time. Part three is the most demanding cognitively as the context becomes yet more disembedded with the examiner leading the candidate into a more thought-provoking discussion based on the candidate's talk in part two. While answering the examiner's questions, the candidate's mind engages in two simultaneous and intersecting processes, one is cognitive and one is linguistic. The cognitive process focuses on the higher-order thinking questions asked by the examiner to try to give appropriate answers, while the linguistic process focuses on producing appropriate language that matches the level of the discussion and that showcases a broad knowledge of general and topic-specific vocabulary to compensate for the lack of contextual clues. The language test task constituted by the IELTS speaking test in this sense is context-reduced and cognitively demanding, whereas the language use task in the classroom is context embedded and arguably undemanding, especially in comparison with the test task. Therefore, the language test task can be located in quadrant D of Cummins's (2000) framework, whereas the classroom teaching task can be located in quadrant A as illustrated in figure (5.1).

Even when a classroom teaching task becomes demanding in situations where the subjects look for specific words to clarify instruction or ideas for their learners, or when they address particular students who need language support, it will still be within an embedded context, and that takes it to quadrant B, but it remains different from the test task, in any way.

Drawing on the discussion presented above about the characteristics of the classroom teaching task and the IELTS speaking test task, it is clear that there is no correspondence between the two tasks in terms of their cognitive demands and the role of context. This raises questions about the cognitive and context validities of the IELTS test task for a classroom teaching task.

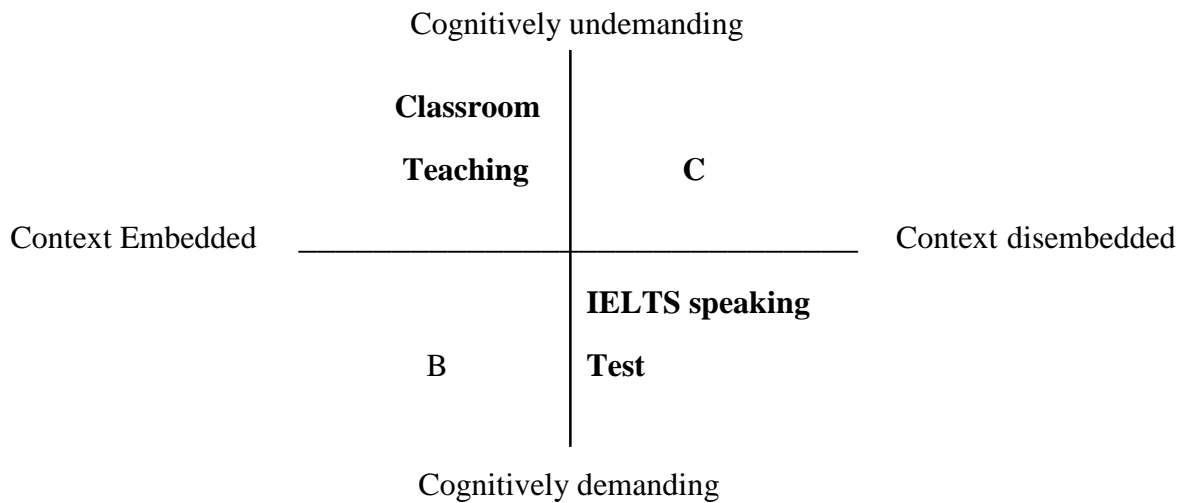


Figure 5.1 *Cognitive demands and context-embeddedness in IELTS and class tasks*

Used with permission of Multilingual Matters

2) The cognitive validity of the IELTS speaking task for a classroom teaching task

Field (2011, p.65) identifies a speaking test as cognitively valid when it engages the candidate in mental processes that are similar to those employed in a speaking activity in the real world. However, the statistical findings presented in chapter 4 and the discussion of the correspondence between the classroom teaching task and the IELTS speaking task presented above demonstrate that there is no similarity. The test task does not engage the same processes as in the classroom. The IELTS test task imposes higher level cognitive demands through a reduced context and time constraint as the candidates try to showcase their lexical diversity. The classroom task, on the other hand, makes less challenging cognitive demands as far as language is concerned due to its context embeddedness and the provision of sufficient preparation time in advance of the task. Though, realistically, a teacher's preparation of a lesson does not necessarily cover all the possible lexical items that can be used in a class, especially in impromptu situations, it can still provide some background knowledge that reduces the cognitive challenges in those situations, unlike the test situation. Moreover, while teaching, the subjects' minds do not, really, engage in the process of vocabulary search to display lexical diversity, but engage in the process of implementing and evaluating teaching strategies that can help their students achieve planned learning objectives. Therefore, we can talk of the classroom teaching task as a cognitively

demanding task only in so far as the teaching skills are concerned, but not the language skills, especially with L2 young learners.

To conclude, the cognitive validity of the IELTS speaking test for the classroom teaching task is apparently weak due to clear differences in the mental processes engaged in each task.

3) The context validity of the IELTS speaking task for a classroom teaching task

Based on the discussion of the correspondence between the IELTS speaking task and the classroom teaching task with reference to Cummins's framework (2000), it is clear that the role of context in each task is different. In the IELTS test task, the role of context is minimized and candidates have to rely on their cognitive and linguistic skills to compensate for the lack of contextual support. That is the reason why they display a wider range of lexical diversity as they answer the interview questions. In the classroom, however, context is maximized through the teacher's preparation of the content and the language to be taught. By using some of Weir's aspects of context validity for speaking (cited in Galaczi and Ffrench, 2011), we can see that there is a mismatch between aspects of the test task context and aspects of the classroom task context. For example, the setting is one of the aspects of context validity, and one of the criteria of this aspect is the purpose of the task. The purpose of the test task is to gauge the candidates' English language proficiency, whereas the purpose of the classroom teaching task is to teach English to young L2 learners. Another aspect is the way the task is administered. The context of the test task in this respect is very different from the classroom teaching task, as the speaking task is conducted through an interview, whereas the classroom teaching task is conducted in different ways including lecturing, questioning, giving instructions, repeating, giving feedback, and so on. The third aspect is the linguistic demands which look totally different. For example, topic familiarity is usually at a low level in the test context except for the first part of the test, but it is usually at a high level in the classroom context due to teacher preparation of the lesson beforehand. Also, lexical resources in the test context are usually used at their optimal potential to compensate for the lack of contextual clues, whereas in the classroom they are used only at the level of the learners and their immediate needs. Finally, the interlocutor's demands are different as well in many ways. For example, student teachers are generally not strongly acquainted with the IELTS examiner, which arguably increases the formality of the candidates' responses and consequently their lexical diversity. In the classroom, student teachers are

generally strongly acquainted with the learners and therefore formality drops along with the lexical diversity which is aimed to match the limited lexical range of the young learners.

The conclusion for this first section on the correspondence between the language use task and the test task in the present study is that there is a lack of correspondence and that this explains why lexical diversity scores on the IELTS speaking test and in classroom teaching do not correlate. The cognitive and context validity of the test task for the classroom teaching task has been discussed and shown to be weak, which will inform the answer to the main question of this study regarding the predictive validity of the IELTS speaking test scores for the performance of student teachers in the classroom.

5.2.2 The correspondence between the characteristics of the teacher trainee and the IELTS candidate

As they call for a correspondence between the test task and the language use task, Bachman and Palmer (1996) recommend a correspondence between the characteristics of the language user and those of the test taker. In the present study, the language users are the subjects as student teachers, and the test takers are the subjects as IELTS candidates. Bachman and Palmer (1996) focus on language ability as the most important characteristic of language proficiency. Though language knowledge seems to be the same construct whether the subjects are trainee teachers or candidates sitting for the IELTS speaking test, it is in reality different when it comes to language use. In a teaching context, the subjects do not use their lexical knowledge at its optimal level as their strategic goal is not to exhibit their lexical diversity, but to use their vocabulary knowledge in the most appropriate way to help their young learners achieve the planned learning outcomes. Therefore, the subjects' language knowledge interacts with their strategic competence in the classroom context to display language ability that matches the level of the learners, and that limits itself to their learning needs. This, partly, answers the question that I asked in section (2.2.3): Do student teachers negotiate social meaning as they interact with a native speaker (the examiner) in the same way as when they interact with young second language learners in a classroom? With reference to Canale and Swain's (1981) framework of communicative competence, and by reference to the clear discrepancies in lexical diversity scores revealed by the analysis in chapter 4, it is clear that the subjects approach each sociocultural context in a different way.

In the case of the present study, student teachers employ different strategies in using lexical items in connection with their grammatical competence as defined by Canale and Swain (1981, p.29). They use a specific lexical diversity range in the classroom which matches the level of their students and their planned objectives. If they fail to do so, they are graded down by their mentors, because that is considered a key characteristic of an efficient teacher in the classroom. That is reflected in the TP competencies under ‘Language and delivery’ (appendix 1). By contrast, on the IELTS test, they try to use a wider range of lexical diversity that can entice the examiner into raising their scores. As far as their sociolinguistic competence is concerned, they use a lexical range that is appropriate for the classroom as a sociocultural context, which includes:

- 1- Using a limited number of key words that match the topic of the lesson and avoiding diversity of lexical items that can confuse learners.
- 2- Performing different types of lexical repeats that are associated with their roles as teachers (giving instruction, asking question and giving feedback).
- 3- Employing specific vocabulary that is associated with the classroom context to control and monitor interaction effectively.

On the IELTS speaking test, student teachers employ different strategies which are related to their sociolinguistic competence. They include:

- 1- Using key words related to the test tasks and any related synonymous words or expressions to demonstrate breadth of vocabulary knowledge, and to make sure they make themselves clear to their audience (the examiner).
- 2- Avoiding repeats of the same lexical items to show lexical diversity, and to avoid redundancy which might affect their score in a negative way.
- 3- Choosing lexical items that demonstrate knowledge of the sociocultural context of the test which include formality and awareness of the status and authority of the examiner and the exam procedures.

It is through a process of accommodation that this happens. In a classroom context, the subjects’ concern is to find an appropriate match between their language knowledge and their communicative goal which is focused on students’ learning. They perform accommodation

through the act of converging with the language proficiency level of their learners. They adapt their speech by using vocabulary that is known by their addressees in order to keep instructions comprehensible and so to achieve the learning goal. They repeat the same lexical items to avoid confusing their addressees when they use alternatives, and to maintain a level of comprehension that helps to achieve planned goals. This has been revealed by the high frequency verbs that relate to classroom instruction like *write, go, come, sit, read*, and so on. (see table 4.8). As indicated by Bell (1984), the addressees, who are the language learners in this case, have a strong impact on the subjects' style, especially in relation to their lexical diversity. Their choice of vocabulary is very much controlled by their knowledge of their students' levels. So, another characteristic of the subjects as language users in a teaching situation is that of the speaker who tries to converge with the level of proficiency of their learners as illustrated in figure (5.2) below:

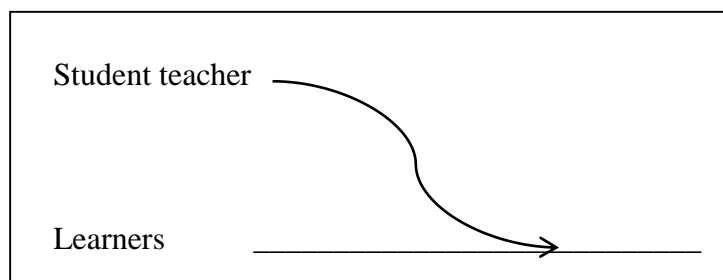


Figure 5.2 *Convergence with the learners' level of language proficiency*

On the speaking test, however, the subjects usually use their lexical diversity to its fullest potential as they are aware that their interlocutor (the interviewer) is assessing their lexical knowledge along with other constructs. The main characteristic of the subjects as IELTS candidates, in this sense, is that of the test taker who tries to showcase their vocabulary knowledge as an important component of their organizational knowledge, and in interaction with their pragmatic knowledge (Bachman & Palmer, 1996). They try to give the interviewer a good impression of their language knowledge. This, in fact, explains why most of the subjects in this research study scored higher in lexical diversity on the test than in the classroom. On the test, as in the classroom, the subjects perform an act of convergence to win the approval of their addressee who is the interviewer, though their act of style shift is totally different. Instead of using high frequency words and repeating them (as they do in the classroom), the subjects use a wide variety of words including the less frequent words to demonstrate a high level of lexical

knowledge, and to try to converge with the examiners' level of proficiency as much as they can in order to win their approval. The illustration below (figure 5.3) demonstrates how the act of convergence which aims at accommodating speech style to the addressee, who is at a higher level of language proficiency and who is assessing the subjects' performance, is reversed compared to the one illustrated in figure (5.2).

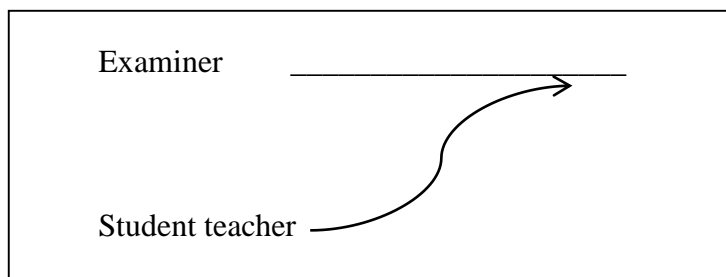


Figure 5.3 *Convergence with the examiner's level of language proficiency*

With reference to Bell's (1984) Style Axiom through which he makes a distinction between intraspeaker and interspeaker variation in language style, it would be appropriate to say that the subjects' awareness of the addressees in each situation makes them pay attention to their speech and to the appropriateness of their lexical choices that make up their lexical diversity. It is this characteristic that makes the difference between the subjects as test takers and as teachers in the classroom.

We can, then, conclude that the correspondence between the characteristics of the subjects as classroom teachers and their characteristics as IELTS test takers is a weak one, which explains why the lexical diversity on the test is clearly different from that in the classroom performance. Therefore, we can say that scores obtained on IELTS are unlikely to be good indicators of what a student teacher can do in the classroom in terms of the type of sociolinguistic language proficiency required by the context. There is more for a teacher to do in the classroom to make themselves understood by their learners than showcasing the breadth of their vocabulary knowledge. Unlike the test situation, teachers in the classroom do not rely solely on language to get their messages through, but use different strategies and techniques that interact with features of classroom interaction including control of interaction, speech modification, elicitation, repair, scaffolding, and repetition. Those features will be discussed in the next

section while considering their impact on the lexical diversity of the subjects as revealed by the qualitative analysis presented in the previous chapter.

5.3 Do student teachers' scores on IELTS correlate with their teaching practicum grades?

In this section, I will discuss the correlation between the subject's scores on the IELTS speaking test as they were rated by their examiners and their grades on TP as they were assigned by their college mentors (MCTs). The statistical results showed that the correlation coefficient is (-.142) and the significance is at (.479). This finding confirms two realities: (1) there is a no relationship between IELTS scores and TP grades, and (2) The predictive validity of IELTS scores for an Education program is weak.

5.3.1 No correlation between IELTS scores and TP grades

One might expect there to be a correlation between a high IELTS score and high TP grades (and conversely between low IELTS scores and low TP grades) on the basis that conscientious students might reasonably be assumed to perform well across the board, and less conscientious students less well. This could be a potential rationale for using IELTS scores in a gate-keeping role for future teachers. This assumption is, however, not borne out by the evidence. The raw scores of the IELTS speaking test and TP grades presented in chapter 4 (table 4.4) show that there is a lack of correspondence between the two sets of scores. The obtained correlation coefficient (-.142) gives clear evidence that there is a no correlation. This confirms the findings of the correlation between the subjects' lexical diversity scores on the test and in the classroom that was discussed in section (4.2.1). In fact, it seems that having a good lexical diversity score on IELTS which means being highly language proficient (in a traditional sense) does not automatically make language students good teachers. However, neither does it prevent them from being good teachers, because different qualities are involved. Trainee teachers need to be able to vary their lexical diversity according to context. What is more, in a number of cases, it was precisely those students who had low D-scores on IELTS had some of the best grades on TP (see T10, T12, T14, T22). Paradoxically perhaps, but tellingly in relation to the validity of using IELTS band 6 as a means to weed out potentially weak teachers, the best teachers are in

many cases the least proficient in English. To understand this finding which agrees with the quantitative and the qualitative discussion of sub-questions 1 and 2 presented previously, we need to start by examining raw scores.

First, it is worth noting here that while the lowest TP score is 6.5 out of a scale of 9 points (with the exception of just one case which scored 5.5), the highest IELTS score is 6.5 (with the exception of one case which scored 7). Two student teachers (T04 & T16) scored 6.5 on both IELTS and TP which means that they scored high on IELTS if we take a norm-referenced approach of ranking students' performance, knowing that 6.5 is above the required band for graduation which is 6. By contrast, the same band puts those two subjects at the lower end of TP ranking. Furthermore, the one teacher (T27) who scored higher than anyone else on the IELTS speaking test (band 7) got 6.5 on TP which indicates that she was not as brilliant in the classroom as her IELTS score indicates.

This can indicate that IELTS scores can be misleading if they are taken just as raw scores with no clarification statement on how they should be interpreted for a teacher education program. In their raw state, those scores leave us with the impression that the subjects' 'high' level of English proficiency will be reflected in their classroom teaching so they get a high TP score as well. However, that IELTS 'high' score does not seem to be enough for a student teacher in the context of the classroom to be assigned a high grade as well. This takes the discussion back to the comment made in section (5.2) where it was noted that there is more for teachers to do in the classroom than just displaying their lexical knowledge, i.e. their language proficiency. I have also highlighted the very significant impact of types of lexical repeats in their reflexive relationship with classroom interaction features and teacher strategies and techniques on the type of language proficiency required of the teacher in the classroom.

Second, based on the tentatively negative correlation between IELTS scores and TP grades, it is worth discussing contradictory scores for some subjects who scored as low as band 5 and 5.5 on the test (below the required band for graduation) but got high grades on TP. In fact, a total of 14 subjects got below band 6 on IELTS, but got 7 or higher on TP. The most palpable discrepancies between scores are recorded by (T21) and (T11) who got band 5.5 and band 5 respectively on IELTS, but got grade 8.5 and 8 on TP, which illustrates the extent to which all the factors that have been discussed in the two previous questions can affect scores and cause

these contradictions to happen. In fact, the differences between the test task and the classroom task, the characteristics of the test taker and the classroom teacher, the addressee design, contextual and cognitive factors, the distinctive classroom context, lexical repeats and their complex relationship with features of classroom interaction, can all explain why performance on the test can look as different and sometimes as contradictory as in the cases of (T21) and (T11). That implies that students who score below band 6 on IELTS speaking test may not be competent speakers of English as per IELTS criteria, but can be excellent teachers who know how to employ their 'limited' linguistic resources in the L2 classroom context in a way that benefits the learners and helps them to achieve the learning objectives. Conversely, a student teacher who scores high on IELTS and is, consequently, labelled a competent user might not have the necessary teaching skills to adapt their language to the requirements of the context of the L2 classroom (e.g. T27), hence may not be a good teacher. In this way, we reach the point where we need to discuss in more detail the predictive validity of IELTS scores for a teacher education program, which is the main argument of this research study.

5.3.2 The predictive validity of IELTS scores for an education program is weak

As seen in the literature review, the predictive validity of a test score is related to its ability to predict subsequent performance. In the case of the present study, the quantitative data analysis in chapter 4 and the discussion of findings presented in this chapter demonstrate that IELTS speaking test scores do not accurately predict performance in the classroom for most of the subjects. If we take band 6 which is the required band for graduation as a criterion for success on IELTS and grade 7 as a criterion for success on TP, we will find that out of the 27 subjects in this study, only 5 subjects (T1, T8, T18, T19, T26) had a good score match (6 or 6.5 on IELTS and 7 or 7.5 on TP) or a reasonable match (6 or 6.5 on IELTS and 8 or 8.5 on TP). The remaining 22 subjects (more than 80%) had mismatching scores that mostly show failure on IELTS (less than band 6), but success on TP (grade 7 or above). Five subjects were below the required IELTS band, but they obtained very high scores on TP (grade 8 and 8.5). This is another piece of evidence that IELTS scores may not be valid predictors of teachers' performance in the classroom. There are different issues that arise here and that can explain why the predictive validity of IELTS scores for an education program is weak:

1) *The issue of language use sampling*

In his review of the development of the IELTS test, Davies (2008) discusses the issue of test sampling. He admits that it is difficult to avoid this problem, because practically speaking you can only test samples but not all instances of language use. He also indicates that it is problematic, because if chosen language samples do not correspond with language use intended for the test taker, issues like the ones identified in the previous chapter and in this one arise. This is indeed what is happening in the case of the present study. The language samples that IELTS suggests for the student teachers do not include any samples from the classroom context. As a result the IELTS scores that students get reflect their proficiency level outside the classroom, whereas the language proficiency that is needed for the classroom depend on some specific factors which are specific to the classroom context, but are not sampled by the IELTS test.

2) *The issue of inappropriateness of inferences*

Messick (1989) indicates that validity is not about the test itself but about inferences based on the test scores. The inferences that are drawn from the subjects' IELTS speaking scores, in the context of the present study, indicate that those who achieve band 6 or higher are 'competent' users of English language, which qualifies them to teach it. The problem lies in the appropriateness of the inference that anyone who achieves band 6 (competent user level) has the ability to teach English, and anyone who does not achieve the band cannot teach English. The data analysis in chapter 4 and the discussion in this chapter demonstrate that those inferences are inappropriate for the context of a Bachelor of Education program. Most of the student teachers who failed to achieve band 6 on the test have been shown by their mentors' grades on TP to be good teachers of English. For college mentors, displaying a good command of the language in the classroom that equals or exceeds IELTS band 6 is considered a quality that can push the subjects' grades higher only in so far as it is well adjusted to the level of the learners and the requirements of the lessons, otherwise it can have the counter effect of decreasing the grade for failure to accommodate speech to the level of young L2 learners. This could be the case for the subject (T27) who scored the highest score on IELTS, but got one of the lowest grades on TP. On the other hand, showing a limited level of language proficiency while displaying ability to accommodate speech to the classroom context in a way that benefits

students would raise the grade for showing abilities to adapt language to teaching and learning in an effective way. This is most probably the case of many subjects, especially T11, T12, T21, T22, and T25.

What college mentors look for is more than just general language proficiency that can be useful to cope with linguistic challenges of theory-based courses offered at the college. As they observe student teachers on teaching practicum, college mentors look for language proficiency which is adapted to the context of the L2 classroom, which uses lexical repeats because they help to achieve the pedagogical focus, and which uses vocabulary and grammar that do not challenge learners beyond their Zone of Proximal development. For mentors, good English language teachers are not those who use their English language to its fullest potential in the classroom, but are those who can manage their linguistic knowledge according to the context where they operate. This gets us to another issue which is the issue of the generalizability of IELTS scores across different contexts.

3) The issue of generalisability of score interpretation

Messick (1989), emphasizes the role of context in evaluating the generalisability of test scores. As discussed earlier in this chapter, the role of context is very important in impacting the speaking performance of teachers, especially in the classroom. The classroom context has been shown to be very different from the test context in a number of different ways that have already been discussed in this chapter. The test context has even been shown to clash with the classroom context in regard with the subjects' roles, the features of interaction, and the audience design. That contradiction has been revealed by the lack correlations and by the qualitative analysis of the classroom interaction and its impact on the teachers' lexical diversity scores.

The IELTS test context does not allow the subjects to be controllers of interaction as in the classroom context due to differences in roles. In the test, the subjects are candidates who cannot ask questions and cannot give feedback. All they can do is to try to answer questions properly. Also, the test context offers them opportunities to showcase their linguistic knowledge at its highest level, whereas the context of the classroom requires that they bring it down to the level of their students through a process of accommodation that functions in the opposite way to that deployed in the test. In addition, the context of the test does not elicit features of interaction that

the classroom context uses, especially with regard to the different categories of repeats and their related types. As a result, it seems that the context of the classroom affects the generalisability of IELTS scores in such a way that they cannot make accurate predictions of teachers' speech.

4) The issue of test misinformed consequences

Messick (1989) links test validity with the purpose it serves and the intended or unintended social consequences. In his unified framework of test validity, he distinguishes between the evidential basis and the consequential basis of test interpretation and test use.

The evidential basis of test interpretation and use is two-fold: (1) construct validity, which is language proficiency in the case of the present study, and (2) the relevance of construct validity to the purpose of the test and its utility, which is its relevance to teaching English in the context of the classroom. Based on the analysis and discussion of the findings, it is clear that the subjects' language proficiency scores on the IELTS speaking test bear little relevance to the classroom context due to lack of correspondence between the test task and the classroom language task, and other factors related to the distinctiveness of the classroom.

The consequential basis of test interpretation and use is two-fold, as well: (1) value implications, which are related to the judgment of a subject's potential value as a future teacher based on the test score, and (2) social consequences, which are related to the subjects' chances to proceed to the final year of the program and graduate, or suspend their registration until they manage to achieve the required IELTS band, or else to give up and decide to opt for another program that does not depend on their IELTS score. In the case of the present study, IELTS band 6 is taken as an indicator of the appropriate language proficiency level that a student teacher should achieve to be able to teach English language. In other words, stakeholders, in the context of the actual study, regard IELTS band 6 as a valid evidence of the construct of English language proficiency (evidential basis) for student teachers to graduate as teachers of English language in UAE schools (consequential basis). However, the discussion in the beginning of this section (5.3) demonstrates that there is a lack of correspondence between the test interpretation based on the scores of the IELTS speaking test and its relevance for the context of language use in the classroom. This finding affects the consequential validity of the IELTS speaking test scores in that the value implication drawn from the test score will not match the social consequences. So,

if IELTS band 6 was found to be of little relevance to the subjects' performance in the classroom, the value judgment by which it qualifies someone to teach English language in UAE school becomes invalid, and the ensuing graduation as a social consequence becomes invalid as well.

5) The issue of using IELTS scores as a 'hard criterion'

While the previous issues seem to be focused on what the IELTS test does not do to simulate the speaking activity that happens in the classroom, it is important to emphasize that it is not the IELTS test itself that is called into question, but the inappropriate use of IELTS test scores for a context that the IELTS test is presumably not designed for. In this respect, I join Rea-Dickins, Kiely and Yu (2011), who found that stakeholders in some universities in the UK 'misuse' IELTS scores by making it a 'hard criterion' for the admission of international students. In the absence of an official document that justifies why IELTS band 6 was chosen as 'a hard criterion' for graduation in the institution where the present study is conducted, I assume that decision makers decided to use IELTS band 6 to ensure that student teachers are competent enough to use English language correctly in the classroom. However, those stakeholders do not seem to have considered the uniqueness of the classroom as a second language use context and its impact on interaction and on teacher talk as discussed earlier in this chapter. Gaps between performance on the test and in the classroom were easily spotted and demonstrated in the quantitative as well as the qualitative analyses. The issue here is in considering the construct of language proficiency as a basic requirement for a teacher in the classroom while ignoring the teaching skills required to make appropriate use of that language proficiency construct.

The third sub-question of this research study is: Do student teachers' scores on IELTS correlate with their teaching practicum grades? The statistical analysis showed that there was no correlation between these scores due to many differences between the two tasks and their contexts discussed throughout this chapter. Moreover, many issues have been identified that contribute to the weak value of the predictive validity of IELTS scores for an education program. Those issues include (1) test sampling which does not include classroom interaction, (2) inappropriate test score-based inferences, (3) lack of generalisability due to contextual factors, (4) misinformed consequences, and (5) misuse of test scores.

5.4 How does classroom interaction affect teachers' lexical diversity scores?

Classroom interaction has some special characteristics that make it different from other kinds of social interaction. The institutional aspect of classroom discourse and its strong dependence on a pedagogical focus impact teacher talk in many ways. Two main and complex factors will be discussed in this section to help answer the second sub-question of this research study: (1) the distinctiveness of the classroom context, and (2) the impact of the different types of lexical repeats.

5.4.1 The distinctive classroom context

In my discussion of the correspondence between the test task and the classroom teaching task in section (5.2.1), I concluded that there seems to be no correspondence, and that the context validity of the IELTS speaking test for classroom teaching is apparently weak. In this section, I will clarify why the classroom context is different and how it impacts the subjects' lexical diversity.

Duranti (1992, p. 80) emphasizes the strong relationship between vocabulary and context, and he indicates that the relationship between them is a dynamic one that changes with the change of the components of the context including the speaker, the hearer, the referents, and the social activities. This claim was clearly confirmed by the quantitative analysis of the subjects' lexical diversity scores in the classroom which showed no correlation with those of the IELTS speaking test. It was also demonstrated by the qualitative analysis of classroom discourse which discussed contextual factors that affect the subjects' lexical diversity in the classroom. In fact, the classroom is a very special context where the relationship between vocabulary and context is governed by the pedagogical focus of most classroom interaction. With an emic perspective informed by the principles of Conversation Analysis, it was possible in the previous chapter to discern the complexity, the dynamic nature, and the uniqueness of the relationship between the subjects' lexical diversity and the context of the classroom.

Drawing on Seedhouse's (2004) 'three way view' of context, we can say that the relationship is a complex one, especially at the level of micro-interaction between (1) the teacher and individual students, (2) the teacher and small groups of students, and (3) the teacher and the class. As many as 16 different types of repeat were identified emanating from those micro-contexts and these

were classified into 6 categories. This provides some evidence of the complexity of the context. Moreover, the way those 16 types interact with features of classroom interaction like elicitation, repair, scaffolding, and so on makes it more complex. Therefore, it would seem difficult -if not impossible- for a speaking test task of about 15 minutes which is conducted solely between the interviewer and the candidate in a non-classroom situation to reflect that complexity and its ensuing impact on the lexical diversity scores of the subjects.

The relationship between lexical diversity and context is also dynamic especially at the level of the L2 classroom as the use of the different types of repetition depends on the pedagogical focus of the teacher. For example, *language drill repeats* would be used frequently in a classroom where a teacher's pedagogical focus is improving students' pronunciation of new words through drills (excerpt 18). However, in another classroom where the teacher's pedagogical focus is developing students' ability to communicate their ideas with adequate support from their teacher, *scaffolding repeats* would be more frequently used (excerpt 20). Yet, some teachers use both types of repeat in one class at different phases of the lesson, depending on the pedagogical focus. The dynamic nature of the classroom context makes lexical diversity vary considerably. So, it would be overgeneralizing to take candidates' lexical diversity scores on the IELTS speaking test as representative of their ability to vary their vocabulary in the classroom.

Finally, the relationship between lexical diversity and the context of the classroom is unique at the institutional level. It is unique especially in the reflexive relationship between the pedagogical focus and the lexical diversity displayed by the teacher. A teacher's lexical diversity will depend very much on the pedagogical focus of the lesson. Teachers cannot use their full range of vocabulary when they are controlled by a specific pedagogical focus. It is more probable that teachers have to bring their lexical potential down to the requirements of the focus. In the same way, the pedagogical focus cannot be met without having teachers use a vocabulary range that is appropriate for the purpose. The different types of repeats and their frequencies discussed in chapter 4 (section 4.4) play a major role in the context of the classroom to help achieve the pedagogical focus. In a reciprocal way, the pedagogical focus dictates the use of some types of repeats that reduce the teachers' lexical diversity scores by comparison with other contexts like the IELTS speaking test.

What we can notice here is that the IELTS test task cannot capture this institutional dimension of the relationship between vocabulary and the pedagogical focus in the classroom context. This explains why there is no correlation between the lexical diversity scores on the IELTS speaking test and the classroom teaching task.

5.4.2 The impact of lexical repeats

As discussed in the literature review on classroom interaction (section 2.4), the classroom is regarded as a unique context with special features that combine together to give us a unique kind of interaction. That combination is a determining factor in making teachers' discourse significantly different from talk in other contexts (Sinclair and Coulthard, 1975; Mehan, 1979; Sinclair & Brazil, 1982; Coulthard, 1985; Seedhouse, 2004; Walsh, 2011 & 2013). Though the features reviewed in section (2.4), namely control of interaction, speech modification, elicitation, repair, scaffolding, and repetition, seem to be equally important in making teachers' discourse a particular one, the findings of the present study, which focuses on the impact of those features on the lexical diversity of the subjects, reveal that repetition is a pivotal characteristic that pervades all the other features. In its interaction with other features and some teacher strategies and techniques, repetition plays an important role in reducing the subjects' lexical diversity scores in the classroom compared to those obtained on the IELTS speaking test. The different categories of repeats and their relevant types (Table 4.10), provide convincing evidence that the subjects' lexical diversity in the classroom context is constrained by contextual factors that the IELTS speaking test task does not replicate. In this section, I will explain how repetitions of particular vocabulary items along with features of classroom interaction and teacher strategies and techniques contribute to giving a lexical range for the subjects that does not correlate with their lexical diversity displayed on the IELTS speaking test.

1) *Control-of-interaction repeats*

In most of the transcripts of the subjects' classes examined in the corpus of classroom data, it was noted that student teachers exert a great deal of control over interaction through a teacher-centred approach. Despite their use of some student-centred activities that give the impression that their pedagogical focus is on developing students' communicative skills and fluency through group and pair work, the subjects usually keep intervening to monitor student-student

interaction in order to keep them oriented to the pedagogical focus. As indicated by Seedhouse (2004), there is a reflexive relationship between interaction and the pedagogical focus. The subjects' control of interaction in this sense is, therefore, very much oriented to their pedagogical focus. Interaction control takes different forms, but the forms that have a direct impact on the subjects' lexical diversity scores are mainly characterized by repeats of their own turns or the students' turns. There are three main types of repeats in this category, (1) *in IRF exchanges*, (2) *in group work*, and (3) *in repair phases* (table 4.11).

In IRF exchanges (Sinclair & Coulthard, 1975), teachers control at least two thirds of the interaction through the first turn where they usually ask questions or give instructions and the third turn where they give feedback. What is noticeable is that this type of exchange is prevalent especially in the first 15-20 minutes of the recorded classes. It makes the class teacher-centred, but it also means that teacher talk is very much concentrated on the pedagogical focus. In fact, during this time the teacher starts with an engaging activity (e.g. a short video, a short chat, a song, or similar activities) that usually aims to activate students' prior knowledge for the main pedagogical focus. After that, the teacher moves to the second stage of the lesson where she introduces the main focus through an IRF exchange that is largely dominated by the teacher who performs many repeats of the main focus (excerpt 1). Consequently, the teacher performs many repeats of the same vocabulary items for a number of turns that allow her to control interaction and to keep her students' attention oriented to the pedagogical focus.

In group work which usually follows the presentation of the pedagogical focus, the subjects assign group activities to their students so they interact in normative orientation to the pedagogical focus. During this stage the subjects control their students' interaction through monitoring group work to make sure that they practise targeted language items in connection with the pedagogical focus. However, what is remarkable in the subjects' turns during monitoring is that they perform a number of repeats of specific words that relate to the pedagogical focus to keep group interaction focused on it (excerpt 2).

In repair phases, the subjects control turn-taking by insisting that students who make mistakes self-correct before giving other students a chance to make corrections (excerpt 3). By performing repeats of specific words, teachers delay other students' turns and exert some control over the sequence organization of classroom interaction.

The question that needs to be asked here is: how much of the teacher control over interaction in the classroom is replicated in the IELTS speaking test to elicit similar lexical diversity? The answer to this question will take us back to Bachman & Palmer's (1996) argument that a test task should correspond to a language use task to expect similar or approximate performance. This feature of interaction control and the associated strategies and techniques employed by the teacher in the classroom teaching task do not seem to be given the right conditions to be applied in the IELTS speaking test. In the test context, the subjects are not given authority over interaction. It is the examiner who controls interaction by deciding when to start and when to finish interaction while being guided by strict IELTS test procedures. Actually, in parts 1 and 2 of the speaking test, we can hardly talk about interaction as the candidates answer the examiner's questions and do not receive any feedback. It is basically a two-dimensional interaction not a three-dimensional interaction as Walsh (2011) depicted classroom interaction. In the third part of the test when the interview becomes more 'interactive' as the examiner starts asking probing questions based on the candidate's answers, this latter is not in control of the interaction. It is the examiner who controls it and who decides how much time is given for the candidate to answer the probing questions and when to interrupt her and take a turn to make a comment or ask another question.

It is clear from this comparison that it is difficult to see any relationship between performing control of interaction in the classroom and on the test. On the contrary, it is easier to see a paradoxical relationship where control of interaction plays an important role in defining the lexical proficiency of the subjects in the classroom, due to different types of repeats; however, on the test, control of interaction is not in the hands of the subjects but it is controlled by the test procedures and the examiner. Therefore, we can say that this very important feature of classroom interaction seems to be deactivated in the IELTS task to the extent that it plays no role in defining the subjects' lexical performance. The asymmetry of participation that Heritage (1997) mentions in his discussion of control of interaction as part of the institutional discourse, and Walsh's (2011) asymmetrical roles of teachers and students seem to be inverted in the test context, but with students replaced by the examiners. The control exerted by the test procedures over interaction is not part of the institutional interaction, hence, it cannot perform the same type of control and it cannot elicit vocabulary in the same way.

2) *Question repeats*

Questions are very important in classroom interaction and especially in teachers' talk. Questions have been researched intensively due to their importance for the classroom context (Long & Sato, 1983; White & Lightbown, 1984; Lightbown & Sapda, 2006, Chaudron, 1988; Ellis, 2008). Heritage (1997) claims that questions give teachers the right to control interaction, and Ellis (2008) finds that the abundance of questions in the classroom is the reason for teachers' control of interaction. White and Lightbown (1984) found that teachers ask up to four questions per minute, and more than 60% of those questions are repetitions of previous questions. In fact, in the present study, question repeats are identified as another distinguishing feature of teacher talk in the classroom and have a direct impact on the lexical diversity scores of the subjects. Like repeats for control of interaction, the category of question repeats in this study has three different types that play a significant role in making teachers' lexical diversity in the classroom different from the IELTS speaking test. The three types are (1) *elicitation and modification repeats*, (2) *strict question-repeats*, and (3) *think-time repeats* (table 11).

In the three types, the common factor that characterizes the subjects' talk is the repetition of questions. *Elicitation* that relies on asking questions is an integral part of teachers' talk and classroom interaction (Walsh, 2011). The analysis of the scripts of the subjects' classes demonstrates that elicitation questions are usually asked at the beginning of the class to try to collect information and language items that have a direct or indirect connection with the pedagogical focus. The elicitations which were examined in the corpus of classroom data are mostly of the type product or process elicitations (Mehan, 1979), whereby the subjects try to elicit factual answers or personal opinions of their students. However, when the subjects do not get immediate responses, they repeat their questions with some *modification* that 'recalibrates' the sequence for the students to help them respond (McLaughlin, 1985). What has been noticed is that the subjects usually modify their elicitation questions by making some changes to their grammatical structures while keeping the same content words that play the role of signposts for the pedagogical focus (excerpt 4 & 5).

Strict question-repeats are a variation of question repeats in non-elicitation phases of the lesson. When the subjects would like to check comprehension or mastery of some language items that have been learnt, they sometimes persist in asking the question by performing strict repeats with

no modification (excerpt 6). In performing question repeats in the same way, the subjects use the same lexical items which reduces their lexical diversity scores.

Think-time repeats follow questions asked with the purpose of giving students some time to process the question and think of an appropriate answer that matches the pedagogical focus of the lesson. As indicated by Walsh (2013, p.35), almost no proper think-time is given to the students to prepare for their answers but the repetition of some questions in the case of the present study is used for that purpose (excerpts 7 & 8). The question repeat in this type is usually identical to the first question as the purpose of the repetition is not to help with the question form or content but with time to think.

Knowing the importance of questions for teacher talk and classroom interaction, and having seen the importance of question repeats in the analysis, it would be legitimate to expect questions to be part of a teacher's speaking test. However, the IELTS speaking test task does not reflect this fundamental characteristic of the language user who asks a large number of questions and makes a considerable number of repeats while using the same vocabulary. Again, with reference to Bachman and Palmer (1996), there is arguably no correspondence between the language test taker who can only answer questions and the language user whose main characteristic is asking and repeating questions.

3) *Feedback repeats*

Feedback is an essential component of the IRF exchange structure (Sinclair & Coulthard, 1975). It allows the teacher to evaluate students' responses in the second turn, and to decide whether to close the exchange or to continue with more turns. Seedhouse (2004, p.186) identifies evaluation as one of three properties of interaction in the L2 classroom where the teacher evaluates the appropriateness of the students' production in orientation to the pedagogical focus announced first by the teacher. The analysis of student teachers' feedback in the present study has found that feedback takes three forms that play an important role in reducing the lexical diversity of the subjects due to the number repeats which are performed. The three types of feedback repeats are (1) *confirmation repeats*, (2) *praise-word repeats*, and (3) *repair repeats* (table 4.10).

Confirmation repeats are frequently performed by the subjects in this study to approve students' answers or contributions, and as a way to keep their young learners motivated for the duration of the activity or the class. It has been noticed that teacher confirmation is often performed through many repeats of key words to show strong approval. For instance, the teacher in excerpt 9 (line 79 to line 94), performs 6 repeats of the word 'green' in just 4 turns (lines 90, 92, 93, and 94) to show her approval of her students' success in showing her things in green in the classroom. It is also noticed that some of those repeats of key words are paired up with praise words which will be discussed next.

Praise-word repeats are found to be the highest in frequency. This is well demonstrated in the findings which revealed that the words *excellent* and *good* are the most frequent ones that the subjects use in the classroom (see 4.3.2). Their impact on the class lexical diversity is a strong one. However, *excellent* does not figure on the IELTS list and *good* appears only as the nineteenth most frequent word. Moreover, it has been found by concordance that *good* is not used in the IELTS test as a praise word, but to describe the quality of things, activities, and so on (appendix 53), which raises questions about the ability of the test task to elicit some of the basic vocabulary for a teacher in the classroom.

Repair repeats take place especially in other-initiated self-repair and in other-initiated other-repair trajectories when the teacher tries to prompt learners who commit errors or their classmates through repetitions of the wrong statement or the wrong word. Lyster and Ranta cited in Lightbown and Spada (2006, p.125) identified six types of corrective feedback including repetition of the wrong utterance produced by the learner while adjusting intonation to draw attention to the error. Seedhouse (2004, p.166) identifies this type of repair as one of eight strategies for conducting repair without 'performing an explicitly expressed unmitigated negative evaluation' in form and accuracy contexts. This is clearly illustrated in excerpt 16 when student teachers start the process of repair by repeating students' utterances verbatim with a rising tone. As a common kind of practice among L2 teachers, this kind of repeat can have a significant impact on their lexical diversity in the classroom.

In its three types identified in this study, feedback as a fundamental component of L2 classroom interaction has a significant impact on the lexical diversity scores of the subjects due to the different lexical repeats they perform. Therefore, to assess the appropriacy of a student's lexical

performance in the classroom, the requirement for feedback which is a key feature of teacher talk, and the repetitions involved need to be taken into consideration. For this reason, it sounds reasonable to say that the IELTS speaking test task does not appear to be designed to capture this key feature, therefore we cannot expect the subjects' scores to reflect it. The issue most probably lies in the reversal of roles in controlling interaction. As mentioned above, in the classroom, the subjects control interaction by asking questions and giving feedback to their students, whereas on the IELTS speaking test they lose that authority over the sequence organization in favour of the examiner and the test procedures. As a result, they take the position of their students in the classroom, yet they do not get any feedback as usually happens in the classroom. So, they are not entitled to give feedback but to answer questions and wait for feedback in the form of a final score after taking the test.

4) *Key-word repeats*

Key words are essential for every class that has a clear pedagogical focus. Whether they are thematic pertaining to a certain topic, or linguistic related to a specific language focus, key words usually pervade teachers' talk as they try to keep learners oriented to the pedagogical focus of the lesson. In the present study, key-word repeats have been identified as important factors in reducing the subjects' lexical diversity in the classroom. Three types have been identified: (1) *lesson key-word repeats*, (2) *activity key-word repeats*, and (3) *story key-word repeats*.

Lesson key-word repeats are words that carry the main theme or the main language focus of the lesson. In all the phases of the class and in most instructions or feedback that the teacher gives to her students those key words appear, which increases their frequency counts and reduces the lexical diversity scores of the subjects. This was clearly illustrated in the previous chapter in excerpts 12, 13 and 14 which are taken from three different phases of the same lesson that had the words 'adjective' and 'noun' as key words. The word 'adjective', for example, was repeated 5 times in excerpt 12 which is taken from the beginning of the class (line 41 to line 59). Then, it was repeated 4 times in excerpt 13 taken from the middle part of the lesson (line 106 to line 111). Finally, it was repeated 5 times in excerpt 14 taken from the last part of the class (line 144 to line 148). By performing such repeats all through the lesson, the subjects keep their students

oriented towards the pedagogical focus of the lesson, but at the same time they reduce their lexical diversity scores considerably.

Activity key-word repeats act similarly to lesson key-word repeats, but at a smaller scale that is restricted to the duration of the activities. Excerpts 15 and 16 are good illustrations of this type of repeat. Though having less impact than lesson key-word repeats, they still represent an important factor that reduces lexical diversity in the classroom.

Story key-word repeats are associated with story-telling or reading classes, especially with young learners. Stories for class reading are usually rhyming stories that rely a lot on the repetition of rhyming words. Consequently, during the reading phases and their related activities like choral reading, the subjects perform many repeats of the same rhyming key words as illustrated in excerpt 17. Story-telling/ reading classes in this way represent a constraining factor for the lexical diversity scores of the subjects.

One might think key-word repeats could be a feature of a candidate's talk in the IELTS speaking test as well, because every question has a specific focus that generates specific key-word repeats. While this sounds reasonable in theory, in reality it is true only to some extent, because in the test the candidates are aware of the necessity to display the breadth of their lexical knowledge, so they would try to use synonyms or alternatives to those key words, whereas in the classroom the subjects act in a different way. In the classroom, they try to repeat the same words even if they know alternatives to avoid confusing their students or diverting their attention away from the pedagogical focus of the lesson.

5) *Approach-related repeats*

This category of repeats is related to strategies that the subjects use following principles of a particular teaching and learning approach. Two main types of repeats have been identified in the data collected. They pertain to two different approaches: (1) *language-drill repeats* pertaining to the audiolingual method, and (2) *scaffolding repeats* pertaining to the communicative language teaching method. Both types of repeats, despite their different theoretical backgrounds, use repeats of words to achieve their goals.

Language-drill repeats are widely practised among the subjects in this study as a way to help their young learners improve pronunciation and memorize new words. The number of repeats can be particularly high when teachers insist that their students pronounce words correctly, or when they use frequent repeats to help memorization. In excerpt 18, the teacher and her students make a total of twenty consecutive repeats of the word ‘eleven’ interrupted once by the teacher to correct pronunciation and two times to encourage students to raise their voices. This number of repeats serves a specific pedagogical focus related to second language learning, but at the same time it contributes to confining the subjects’ lexical diversity to a narrow range.

Scaffolding repeats are also widely used by the subjects in this study to help their students develop their communicative ability. As suggested by Bruner (1990), subjects support their learners with the vocabulary that is needed until they see them internalizing it and using it properly. This is demonstrated in excerpt 20, where the teacher uses modelling in order to guide the students into producing the target form (Walsh 2013, p.84). In modelling and in providing guidance the subjects perform some repeats though not in the same fashion as in language-drill repeats, but these repeats have some impact on their lexical diversity scores.

Approach-related repeats are another feature that characterizes student teachers’ talk in the classroom, and their impact on the lexical diversity of the subjects is evident. For college and school mentors, approach-related repeats indicate student teachers’ awareness of effective strategies that work better for their students’ learning. Being able to perform those numbers of repeats appropriately to facilitate learning qualifies the subjects as effective teachers according to the teaching practicum requirements. However, this important feature of a teacher’s task in the classroom is not displayed in the IELTS speaking test task because the subjects are given questions to answer not a lesson to teach. Consequently, we can say that it is not within the scope of the IELTS speaking test to consider this characteristic which is specific to the language classroom context.

6) *Procedural repeats*

Procedural repeats are an intrinsic part of any classroom, especially for young learners. Teachers need to make themselves clear about what learners need to do and how they should behave. They keep monitoring students’ behaviour and participation in class activities through clear

instructions and repeats of specific lexical items according to the planned tasks and the needs of their students. Two types of procedural repeats have been identified in the data collected for this study: (1) *instruction-clarification repeats*, and (2) *classroom-management repeats*.

Instruction-clarification repeats are used by teachers as they give instructions for their students to carry out assigned activities in order to achieve the planned objectives. Some lexical items are repeated more frequently than others due to their importance for the students to understand clearly what is required from them. Excerpt 22 is a good example of procedural repeats which are used to make sure that the instructions are clear enough before starting an activity. The teacher in this excerpt takes fourteen successive turns (lines 47 to 60) to explain instruction clearly for her students. In the meantime, she performs multiple repeats of different lexical items among which the word 'student' gains the highest number of repeats, which indicates the pedagogical focus of the teacher who wants the activity to be student-centred. Similar types of repeats have also been noticed when conducting instruction clarification for students or groups who show lack of understanding while working on an assigned task. All of those repeats weigh heavily on the lexical scores of the subjects in the L2 classroom.

Classroom-management repeats are more concerned with managing students' behaviour so lessons run smoothly and achieve their learning aims. Different types of classroom management repeats have been identified in the analysis which focus on regulating students' behaviour at the beginning, during, and at the end of conducted activities. Excerpts 25, 26, 27, and 28 cover different instances in a class when the teacher resorts to repeats of lexical items that are designed to help her manage her students' behaviour while keeping them oriented to the pedagogical focus of the lesson. Those management repeats are characteristic of the classroom as a social setting. Their frequency is part of a teacher's success in managing teaching and learning effectively. The number of lexical repeats they require makes them effective but at the same time they contribute to reducing the subjects' lexical diversity scores.

Procedural repeats are another key feature of teacher talk in the classroom. Their role in organizing learning is crucial for the success of the learning process in the L2 classroom. Student teachers perform procedural repeats many times during their classes to make sure their students know what to do and how to behave properly in order to learn effectively. When evaluated by their college mentors, student teachers are assessed for their abilities to give clear instructions

and to clarify them when necessary, and for their success in demonstrating effective classroom management. The impact of those repeats on the subjects' lexical diversity is negative as far as their scores are compared with the IELTS speaking test scores, but it could be positive as far as their teaching practicum grades are concerned –though teacher repeats can represent only a small component of one of the 5 competencies that MCTs consider in grading student teachers on TP. This paradox will be discussed in more detail in the next section on the correlation between IELTS speaking test scores and college-mentors' grades of the subjects' performance on the teaching practicum.

In conclusion, this section has attempted to answer the second sub-question of this research study: how does classroom interaction affect teachers' lexical diversity scores in the classroom? The evidence adduced above indicates that repetition of lexical items is a major characteristic of classroom interaction and teacher talk. It plays a significant role in making student teachers' lexical diversity scores in the classroom very different from their lexical diversity on the IELTS speaking test. The classroom as a distinctive context generates language use tasks and elicits teacher characteristics that the IELTS speaking test context does not seem to capture. 'The architecture of classroom interaction' (Seedhouse, 2004) with its complex relationship between context and the pedagogical focus of the teacher creates distinctive features of interaction and types of repeats of lexical items that impact on the teacher's talk in a significant way. The 16 types of types that have been identified in this research study represent the cornerstone of the L2 classroom interaction that is conducive to effective learning, but at the same time they are very influential in reducing teachers' lexical diversity scores. Their association with basic features of classroom interaction and teacher strategies and techniques makes them important in any L2 classroom. Moreover, their appropriate use by student teachers on teaching practicum is regarded as a good practice that raises their teaching practicum grades and gets them credits that qualify them as good teachers. However, those same repeats lower lexical diversity scores considerably due to their frequency, which creates a gap between the range of scores obtained in the classroom and the one obtained on the IELTS speaking test.

5.5 Conclusion

In this chapter the main question of this research study on the predictive validity of IELTS scores for a teacher program in the UAE was addressed by drawing together the implications of the results presented in Chapter 4 in relation to the three sub-questions proposed as research questions in Chapter 1, that is to say:

- 1) Do student teachers display similar lexical diversity on the IELTS speaking test and in the classroom?
- 2) Do IELTS scores match college mentors' grades for student teachers on teaching practicum?
- 3) How does classroom interaction affect student teachers' lexical diversity scores?

The first question was answered by referring to the statistics that demonstrate that student teachers do not display the same lexical diversity on the IELTS speaking test as in the classroom. The comparison between D-scores, the means, and the standard deviations demonstrate that they are different. The correlation coefficient test confirms the differences by revealing no correlation. This finding was supported by the qualitative analysis which demonstrated that there are not only differences but also contradictions between the IELTS test task and the classroom teaching task and between the characteristics of the subjects as test takers and as student teachers. Lack of correspondence was also demonstrated in the cognitive demands of each task and the contextual support. A contradiction was also highlighted in discussing style shift and addressee design while taking the test and in the classroom.

The second question was answered by demonstrating the distinctive nature of the classroom context which involves complexity, dynamism, and uniqueness in the relationship between lexical diversity and the pedagogical focus. It was also answered through a qualitative analysis and a discussion of the impact of the types of repeats that teachers perform in a classroom in their interaction with features of classroom interaction and teacher strategies and techniques.

Finally, the third question was answered through a discussion of the statistical findings which came identical to the findings of the first question with a weak and a negative correlation between IELTS scores and TP grades. It was supported by a discussion of issues related to test

sampling, score interpretation, score generalisability, social consequences, and score misuse that impact the predictive validity of those scores.

The sum of the three answers give a clear answer to the main question of the study, which is that IELTS scores are not good predictors of student teachers' scores on teaching practicum. Moreover, the qualitative analysis of teacher talk in the classroom make us conclude that IELTS scores may not be good predictors of student teachers' performance in the classroom, due mainly to clear differences between the test task and the classroom teaching task, the distinctiveness of the classroom context, and the resulting test issues.

Chapter 6: Conclusion

6.1 Introduction

This chapter brings this research study to a close where implications for practice will be discussed with a focus on language proficiency testing in teacher education. Then, contributions to debates in applied linguistics will be summarized. They will cover three different areas and disciplines, including (1) predictive validity and language testing, (2) lexical diversity, lexical proficiency and language proficiency, and (3) teacher repeats and classroom interaction. Contributions will also cover corpus linguistics and the L2 classroom discourse. Finally, this chapter will draw attention to some of the limitations of this research study for consideration in future research studies.

6.2 Implications for practice

By looking back at the analysis and the discussion chapters, the answers for the research sub-questions indicate that there are some clear issues associated with using IELTS scores as predictors of student teachers' subsequent performance in the teaching context. Therefore, using those scores as a hard criterion to grant or deny student teachers access to the final year of the B.Ed. program in order to graduate as teachers of English in the UAE schools seems to be a flawed practice that needs to be revisited in the light of the findings of the present study. While being critical of using IELTS scores in a rigid way and with no validating argument, I reiterate my point of view that the problem does not lie in the IELTS test itself but in the way scores are interpreted and used for high stake decisions. Based on this and on the findings of this research study, I suggest an alternative way of using IELTS scores as predictors of student teachers' performance in the classroom in a way that bridges the gap between the test and the classroom context. This alternative way consists of two main steps: (1) developing a complementary assessment to the IELTS test that assesses classroom-based English language proficiency (henceforth CBELP), and (2) developing an argument that validates interpretations assigned to IELTS and CBELP scores and to ensuing decisions. So, what is meant by complementary assessment? And how can it capture what IELTS test cannot capture in the classroom context?

6.2.1 A complementary assessment for classroom-based language proficiency

A complementary assessment in the case of the present study refers to an assessment tool that can do what the IELTS test has been proven to be unable to do in order to gauge student teachers' language proficiency in the classroom, like using test tasks that correspond to classroom teaching tasks, or simulating classroom interaction with its very distinctive features. Therefore, the only plausible and feasible way, in my opinion, is through assessing student teachers' language proficiency while teaching, and assigning them band scores that can be used along with IELTS test scores to represent their language proficiency (this will be explained in detail in the next section on the validation-argument). Now, how can this complementary form of assessment (CBELP assessment) be conceived and made operational to gauge student teachers' English language proficiency in the classroom? And how can its scores be used along with the IELTS test scores?

1) The design of the CBELP assessment

To answer the first question, we need to draw on Freeman et al. (2015) model of 'English-for-Teaching' to design a rubric that can be used by college mentors to rate student teachers' language proficiency in the classroom while observing them on TP. The rubric will use the three main functional areas of classroom tasks and routines performed by teachers, and identified by Freeman et al. (2015, p.134) as the most common areas where English language plays the dual role of being part of the curriculum content and the means through which that content is delivered. As illustrated in figure (2.3) the three functional areas are: (1) managing the classroom, (2) understanding and communicating lesson content, and (3) assessing students and giving feedback. Those three areas will be mapped onto the four speaking sub-skills of the IELTS test, namely *fluency and coherence*, *lexical resources*, *grammatical range and accuracy*, and *pronunciation*. Besides, four levels of classroom language proficiency equivalent to bands 5, 6, 7 and 8 of the IELTS test will be used to score performance, and like the IELTS test, scores will be reported in whole and half bands. The rationale behind this scale is that B.Ed. students enter the program at band 5 and some of them –very few though- achieve band 8 in speaking as they graduate. However, the descriptors will not be graded across the different bands in the same way as in the IELTS test. Rather the opposite in some cases, because the classroom context requires less diversity and complexity and more repetition as we go higher in the scale to ensure

effective teaching and learning. So, those who use high-frequency words for classroom management, for example, are more likely to achieve a higher band than those who use low-frequency or advanced vocabulary. This has already been demonstrated by the quantitative analysis which showed that some of the subjects who got high lexical diversity scores were among those who got the lowest mentor grades on TP (see the cases of T04, T07, T15, T17, & T27 in table 4.4). This was also discussed at length in chapter 5 (refer to sections 5.2.1 and 5.4.1).

2) Using the CBELP rubric

A model of the rubric is suggested in table (6.1) below, where generic descriptors have been developed based on no specific curriculum, but as a general model.

To ensure standardization of practices and an optimum level of objectivity in grading student teachers' performance, training and moderation sessions using TP video-recorded classes will be scheduled for the B.Ed. faculty across the system of the institution. Furthermore, to secure more objectivity of scoring, the CBELP assessment will be assigned to a different mentor than the one who supervises the student teacher on TP.

It is worth mentioning here that this generic rubric can be made specific by specialists in assessment from the B.Ed. program who have adequate knowledge of the UAE national curriculum and international school curricula where student teachers have their TP placement.

6.2.2 Establishing the predictive validity of the IELTS test and the CBELP assessment for a teacher education program

The use of the CBELP assessment rubric illustrated above or an adapted version of it will help to eliminate all the validity issues associated with the use of the IELTS test scores to predict student teachers' subsequent performance (see chapter 5, section 5.4.2) and will, therefore, help to establish the predictive validity of IELTS scores. Lack of correspondence between the IELTS test task and the classroom teaching task is a major cause for most of the identified issues. Differences between the two tasks have been discussed in detail and they have been found to be related mainly to the volume of cognitive demands and contextual support, the ways strategic competence is deployed, the types of speech accommodation strategies, the impact of the

distinctiveness of the classroom context, the impact of teacher repeats in conjunction with features of classroom interaction.

Table 6.1 *An assessment rubric for classroom-based language proficiency*

		Fluency and coherence	Lexical resources	Grammatical range and accuracy	Pronunciation
8	<i>Classroom management</i>	<ul style="list-style-type: none"> speaks fluently while adjusting the pace to the level of the learners 	<ul style="list-style-type: none"> uses an adequate range of vocabulary items uses high frequency words all the time uses appropriate vocabulary for the level of the learners 	<ul style="list-style-type: none"> uses an adequate range of simple and complex structures produces accurate sentences and word forms uses complex sentences which are appropriate for the level of the learners 	<ul style="list-style-type: none"> uses a wide range of pronunciation features makes speech clear enough to all learners Models correct pronunciation through clear articulation of sounds
	<i>Content understanding and delivery</i>	<ul style="list-style-type: none"> speaks coherently all the time uses repetition adequately 			
	<i>Assessment and feedback</i>				
7	<i>Classroom management</i>	<ul style="list-style-type: none"> speaks fluently while trying to adjust the pace to the level of the learners 	<ul style="list-style-type: none"> uses a reasonable range of vocabulary items uses high frequency words most of the time uses mostly appropriate vocabulary for the level of the learners 	<ul style="list-style-type: none"> uses a reasonable range of simple and complex structures produces accurate sentences and word forms most of the time uses complex sentences which are mostly appropriate for the level of the learners 	<ul style="list-style-type: none"> uses a reasonable range of pronunciation features makes speech clear most of the time Tries to model correct pronunciation through clear articulation of sounds
	<i>Content understanding & delivery</i>	<ul style="list-style-type: none"> speaks coherently most of the time uses repetition adequately 			
	<i>Assessment & feedback</i>				
6	<i>Classroom management</i>	<ul style="list-style-type: none"> shows some hesitancy and some difficulty to adjust the pace to the level of the learners/ speaks fluently while showing difficulty to adjust pace to the level of the learners 	<ul style="list-style-type: none"> uses a range of vocabulary items uses a few low frequency words uses some inappropriate vocabulary for the level of the learners 	<ul style="list-style-type: none"> uses a range of simple and complex structures produces some inaccurate sentences and word forms uses some complex sentences that can be confusing for the learners 	<ul style="list-style-type: none"> uses a limited range of pronunciation features Faces some difficulties to make speech clear for the learners Faces difficulties to model correct pronunciation of some sounds
	<i>Content understanding & delivery</i>	<ul style="list-style-type: none"> tries to speak coherently 			
	<i>Assessment & feedback</i>	<ul style="list-style-type: none"> does not use repetition adequately 			
5	<i>Classroom management</i>	<ul style="list-style-type: none"> shows frequent hesitancy that affects fluency and message clarity/ speaks fluently but fails to adjust pace to the level of the learners 	<ul style="list-style-type: none"> uses a wide range of vocabulary items uses low frequency words uses inappropriate vocabulary for the level of the learners 	<ul style="list-style-type: none"> uses a wide range of simple and complex structures produces frequent inaccuracies in sentence and word forms uses complex sentences and structures that confuse the learners 	<ul style="list-style-type: none"> uses a very limited range of pronunciation features Faces difficulties to make speech clear for the learners Fails to model correct pronunciation of sounds
	<i>Content understanding & delivery</i>	<ul style="list-style-type: none"> produces incoherent utterances 			
	<i>Assessment & feedback</i>	<ul style="list-style-type: none"> does not use repetition 			

1) *Eliminating IELTS scores validity issues*

Five validity issues have been identified in the use of IELTS scores to predict student teachers' subsequent performance. The complementary assessment (CBELP) suggested in this research study with its focus on assessing student teachers' language proficiency as they practice teaching in the classroom will help to eliminate the identified issues in order for the predictive validity of the IELTS test scores for a teacher education program to be established.

- The first issue of *language sampling* in the IELTS test will be resolved by the fact that the complementary assessment is conducted in the classroom which makes it an authentic sample of language use that validates assigned scores as indicators of language proficiency.
- The second issue of *inappropriateness of inferences* based on the IELTS test scores will also be resolved by the fact that obtained scores on the complementary assessment are based on direct observation of the student teachers' performance in an authentic situation of language use which is the classroom. Consequently, any score interpretation and related inferences are based on a directly observed performance that reflects student teachers' language proficiency in the classroom.
- The third issue of *the generalizability of score interpretation* will be eliminated as a result of resolving the first two issues. Basing scores on an authentic performance sample and making appropriate inferences makes generalizability of score interpretation very possible, because it can be applied to all other teaching performances since the context is the same which is the L2 classroom, and the interaction is the same with the teacher controlling most of it.
- The fourth issue of *misinformed consequences* is resolved by the fact that the CBELP assessment is based on a classroom teaching performance. Both the evidential and the consequential bases of test score interpretation are validated. The evidential basis is validated through the assessment of the construct of classroom-based language proficiency by observing the subjects while teaching as an evidence that can predict subsequent teaching performances. The consequential basis is validated through making a judgment of the subjects' potential as teachers of English based on assessing them in the classroom which is the context where they will be performing English language teaching after graduation. This, as a consequence, validates the high-stake score-based decisions of granting or denying them access to the final year of the B.Ed. program before graduation.

- Finally, the issue of *using IELTS score as a hard criterion* will be eliminated by the fact that the CBELP assessment score will be considered along with the IELTS score in making inferences about student teachers' performances and making subsequent decisions.

2) Developing an argument that validates IELTS and CBELP score interpretation and ensuing decisions

In his article on validating score interpretation and uses, Kane (2011, p. 8) says 'The interpretive argument specifies the reasoning involved in using the assessment scores to draw conclusions and make decisions and is applied every time the assessment scores are used.' He also says 'a proposed interpretation or use is clearly justified, because its inferences and assumptions are supported by empirical evidence and/ or are highly plausible *a priori*' (2011, p.10).

In the absence of an official statement which gives a clear interpretive argument explaining how IELTS test scores are validated to make inferences and high-stake decisions about the future of the B.Ed. students, Kane's (1992) argument-based validation constitutes a useful frame of reference along with the arguments presented in chapter 5 to develop an argument-based validation for the use of the CBELP scores in conjunction with IELTS scores. The developed argument will establish validity through (1) giving a clear explanation of the way IELTS and CBELP scores will be interpreted and (2) justifying score-based decisions.

As noted before, the main problem that is causing validity issues in using IELTS scores to predict student teachers' subsequent performance in the classroom is the lack of correspondence between the test task and the classroom task. The suggested CBELP assessment is meant to establish that missing correspondence to be able to eliminate the identified issues and to provide a plausible interpretive argument (Kane, 1992) that can justify score interpretation and subsequent decisions. So how can correspondence be established between two tasks that seem to be standing at two opposite ends?

- a) Establishing links between the IELTS test task and classroom teaching assessment to validate score interpretation

First of all, my argument in suggesting a complementary assessment and not an alternative assessment to IELTS is built on an assumption that I made in the absence of an official statement

by the institution where this research study has been conducted, and based on my experience as a B.Ed. faculty in this institution for more than ten years. So, if we go by the assumption that policy makers, in the context of the present study, chose IELTS as an internationally renowned language proficiency test as a way to certify the quality of the awarded degree, or to gain some external validation in order to get easier access to the job market, especially that most of the private schools in the UAE are international schools, the suggested complementary assessment can get the degree internal validation by which English language proficiency which is used in the L2 classroom is properly assessed and attested. This policy is, in reality, followed by many other institutions in the UAE which require a specific IELTS band score along with the degree to grant applicants a job. It has been implemented by Abu Dhabi Education Council, and recently by the UAE Ministry of Education who require, beside the teaching degree, an IELTS report attesting that the candidate achieved band 7 as a minimum requirement for an application to be screened and processed, then for a candidate to be interviewed. Therefore, to suggest discarding IELTS test score and replacing it with the suggested CBELP assessment would make the suggestion implausible, and unpractical, at least at this stage.

Another good reason for me to suggest CBELP as a complementary assessment to IELTS and not as an alternative is that despite its virtues, especially in connecting IELTS scores with classroom practice, it does not assess academic English that IELTS tests, and that remains necessary for teachers to do their job properly. In fact, a teacher's language proficiency is not, or rather should not be confined within the walls of the classroom but extends to other academic contexts where classroom-based language proficiency would not be enough to cope with the requirements of communication. What I mean here, is that teachers engage in various types of communications outside the classroom where they need to use their general and academic English language proficiency. They attend regular professional development events, some of which are mandatory and some are voluntary. Some teachers deliver professional development sessions and engage in onsite and online debates with their colleagues over teaching and learning topics and issues. With the extraordinary development in social media and the myriad of educational software, applications and websites, they need in the first place their academic and general English language proficiency to be able to communicate with other teachers and professionals all over the world and to use all the available resources in order to keep abreast of the all the developments in the field of education. Moreover, some of the teachers continue their

studies in education while performing their job of teaching. For all these activities and much more that are not within the scope of the present study to list, graduating teachers do need to demonstrate a minimum level of academic language proficiency that can be measured by the IELTS test along with their classroom-based language proficiency that can be gauged by the suggested complementary assessment (CBELP).

b) Validating interpretation of IELTS and CBELP scores and related decisions

As for score interpretation, the CBELP assessment will be used along with IELTS test scores to make interpretations, inferences and decisions. However, to avoid obscuring interpretations, averaging the two scores is not suggested as an option here unless both scores meet the minimum required band, because the average band of two extreme scores may not be a true or an approximate reflection of any of the extreme scores. In other words, if a student teacher gets band 5.5 on the IELTS test (which is the most frequent score band obtained by the subjects in the present study) and 7.5 on CBELP assessment, the difference will be two bands, and the average score will be 6.5. This means the average score will be at a whole band distance from each score, which can mislead interpretations and misinform decisions. Hence, validity issues will persist.

To validate score interpretation, it would be appropriate to accept any two scores within one-band difference or less as indicators of academic and class-based language proficiency. For example, if a student teacher gets band 6 on the IELTS test and band 7 on the CBELP assessment or vice-versa, the interpretation is that her language proficiency on the IELTS test approximates her language proficiency in the classroom. In this case, if the two scores meet the minimum required band, the student's language proficiency is considered good enough, and the ensuing decision is to grant her access to the final year for graduation as a teacher of English in the UAE schools.

On the other hand, if the two scores are at more than a band difference, the interpretation should be that one of them needs to be improved to achieve the same or an approximate level as the other one. For instance, if a student teacher gets band 7 on IELTS and 5.5 on the CBELP assessment, the score-interpretation is that the student is a good user of English language, but needs to work on and develop classroom-based language proficiency. The decision that should

follow is a recommendation for the B.Ed. department to provide classroom-based language support for that student before scheduling another classroom observation for the CBELP assessment. In this case, the student will not be allowed to register in year 4 until she achieves the required band for CBELP which is band 7. Instead, she can be asked to take the Teaching Practicum course again to improve her classroom-based language proficiency, if the given classroom-based language support proved to be insufficient. In the opposite case, i.e., those students who get low scores on the IELTS test, and good or high scores on CBELP should be given support in their English language skills before they are granted access to year 4. However, by looking at the CBELP rubric, it looks unlikely that a student teacher with low academic English language proficiency would achieve band 7 or 8 on the CBELP assessment, because they will not be able to demonstrate fluency, coherence, and especially grammatical accuracy in the classroom at that level. Now, the question is: what is the minimum band score for the IELTS test and the CBELP assessment below which a student cannot be granted access to the final year of the B.Ed. program, and on which basis is it defined?

c) Validating the minimum band score

While making a suggestion to base score interpretation and ensuing decisions on both the IELTS test score and the CBELP assessment score to avoid the effect of the ‘hard criterion’, setting a minimum band score for each component is a necessity to secure a minimum level of academic and classroom-based language proficiency that would help graduates to do their job properly. Therefore, keeping the same required score of the IELTS test which is band 6 seems reasonable to me, because that band score is equivalent to a ‘competent user’ of English according to the developers of the test, which is enough to do all the ‘para-teaching’ tasks and communications. However, for CBELP assessment, I would recommend band 7 as a minimum score, for two main reasons: (1) classroom-based language proficiency should be at a ‘good level’ (in IELTS terms) that is even higher than academic language proficiency, to ensure effective teaching and learning in the L2 classroom which is the first target of the awarded degree, and (2) by looking at the suggested rubric, band 7 descriptors indicate that a student teacher displays all the expected CBELP skills, whereas band 6 descriptors indicate that the student teacher fails to display some of them. In this way, it is necessary for student teachers in the B.Ed. program to

get at least band 6 on the IELTS test and band 7 on the CBELP assessment to be qualified as teachers of English in the UAE.

6.3 Contributions to the current state of knowledge

The contribution of this research study to the debate in the literature on applied linguistics is manifold. First, the predictive validity of IELTS scores for the qualification of student teachers as teachers of English in the UAE has been challenged and has been found weak due to disparities between the test as a measurement tool and the context of language use which is the classroom. Second, lexical proficiency is found to be a variable construct that depends on the context of language use, therefore it should not be taken as an indicator of language proficiency in a broad sense. Third, a taxonomy of teacher repeats of lexical items has been created to reflect one of the most salient characteristics of teacher talk that impact lexical proficiency in the classroom. Last but not the least, a corpus of the L2 classroom interaction and student teachers' performance on the IELTS 'mock' speaking test is gleaned to provide empirical data for the present study and for further investigations in relevant areas.

6.3.1 The predictive validity of IELTS for a teacher education program

As revealed by the literature review in chapter 2, research studies on the IELTS predictive validity have not been conclusive due to conflicting findings. Most of those studies were conducted in tertiary education institutions where correlations between students' pre-admission and post-admission IELTS scores were conducted to gauge the predictive validity of IELTS. However, those studies had the limitation of studying *truncated* samples (Alderson et al., 1995), because those who fail the pre-entry test are automatically excluded from the admitted sample on which studies are usually conducted. Except for Rumsey's study (2013) which investigated CEPA predictive validity (a similar test to IELTS) in the UAE while using the whole sample, I believe that my research study is the first one to research IELTS predictive validity while using the whole sample of subjects. That was possible because the subjects took the test as a pre-graduation requirement not as a pre-entry requirement. So, even when students do not achieve the required band, they stay in the program and they re-sit the test until they achieve band 6.

Among all the reviewed research studies on IELTS predictive validity, only one study (Elder, 1993) was conducted for a teacher education program, but it investigated the predictive validity of IELTS as an entry requirement. So, to the best of my knowledge, my research study is the first one to investigate the predictive validity of IELTS as a program-exit requirement. Elder's (1993) findings showed that IELTS predictive validity as an indicator of success in a teacher education program was weak, and she concluded that language proficiency is a weak predictor of success in an education program, because many other factors related to education come into play, like 'subject knowledge, scholastic aptitude, cultural adaptability, understanding of classroom role relationships, motivation, interactive style' (Elder 1993, p.87). In fact, the findings of the present research study come to confirm and to complement what Elder (1993) found. IELTS scores are found to be weak predictors of performance on TP, mainly due to factors related to the classroom context which has distinctive features that the IELTS test does not capture and does not replicate.

The use of IELTS scores as a 'hard criterion' (Rea-Dickins, Kiely & Yu, 2011) to decide on the official qualification of the B. Ed. students as teachers of English in the UAE is found to be an inappropriate practice. The collected data revealed that student teachers' lexical diversity and scores on the IELTS speaking test had little relevance to their lexical diversity and grades on TP in schools. Therefore, this finding confirms Elder's finding that language proficiency for teachers is more than just performance on a standardized test but it involves all the factors that make up the L2 classroom interaction as discussed in previous chapters.

6.3.2 The variability of lexical diversity according to the context

The second contribution is made to the debate on the viability of lexical diversity and lexical proficiency as indicators of lexical proficiency. As reviewed in chapter 2 in section (2.3) and sub-section (2.3.1), lexical studies have argued for the strength of lexical diversity and lexical proficiency as indicators of language proficiency (Daller, Van Hout & Treffers-Daller, 2003; Laufer & Goldstein, 2004; Alderson, 2005; Zareva, Schwanenflugel & Nikolova, 2005; Read & Nation, 2006; Albrechtsen, Hasstrup & Henrikson, 2008; Crossley et al., 2011; Crossley, Salsbury & Mcnamara, 2012). However, the actual study demonstrates that lexical diversity and lexical proficiency should not be taken as viable indicators of language proficiency in all the contexts of language use. The L2 classroom has been identified as a distinctive context with a

special interactional architecture that elicits a different lexical profile than the one displayed on the IELTS speaking test. The subjects in the present study display a ‘lower’ profile of lexical proficiency in the classroom that is mainly meant to cope with the requirements of interaction in the L2 classroom with its distinguished features and roles of participants that have been discussed in chapter 5. In fact, displaying a lower lexical profile in the classroom to match the level of the learners gains the approval of the MCT who observes the student teacher, and as a result the language proficiency assessment score goes higher. This is already suggested in the CBELP rubric (table 6.1) which assigns a higher score (band 8) for those student teachers who ‘use high frequency words all the time’, and a lower score (band 5) for those who ‘use low frequency words’. In contrast with general language proficiency tests (like IELTS) where lexical proficiency scores have been proven to rise as test scores rise, this study demonstrates that lexical proficiency scores go in the opposite direction in the context of the classroom, i.e., they fall as language proficiency scores go higher, and vice-versa. Therefore, taking lexical proficiency as a viable indicator of language proficiency, should not be generalised to all contexts, but should remain a context-bound finding.

6.3.3 A taxonomy of teacher repeats in the L2 classroom

Repetition has already been identified as a key feature of teacher talk in the L2 classroom. The reviewed literature focused on some general and specific functions of teacher repeats like acting as ‘question rejoinders’ (Halliday & Hassan, 1976), persisting in getting answers for questions (White & Lightbown, 1984), conducting repair (Seedhouse, 2004), facilitating acquisition (Fillmore in Ellis, 2008), and asking for elaboration, confirming a response, or scaffolding (Park, 2014). However, by focusing on the impact of repeats on the lexical diversity of the subjects in this research study, 16 specific types of repeats have been identified with different functions for teaching and learning, and varying impact on the teacher’s lexical diversity in the classroom. Few of the identified repeat types are closely associated with the functions already mentioned in the reviewed literature such as, repeats for repair, repeats to confirm responses, and repeats for scaffolding, but most of the others like repeats in IRF exchanges, lesson keyword repeats, language-drill repeats, and others are -to the best of my knowledge- identified for the first time in the literature on classroom interaction and teacher talk. To make it easier to

distinguish between the 16 types and their uses in a lesson, they have been classified into 6 main categories that are outlined in (table) 6.2 below.

While identifying those 16 types of repeats represents one of the significant achievements of this research study in terms of highlighting a major feature of teacher talk that makes language proficiency in the classroom look different than on a test, I am not claiming that my taxonomy covers all the possible types because I am quite sure that with more data collected from different contexts, it would be possible to identify more types. By different contexts, I mean different types of schools, because the 27 classrooms from which data has been collected and analysed in this study are all government schools in the UAE.

Table 6.2 *An outline of the taxonomy of teacher repeats*

	Category	Type
1	Interaction-control repeats	In IRF exchanges
		In group work
		In repair phases
2	Question repeats	Elicitation and modification repeats
		Strict question-repeats
		Think-time repeats
3	Feedback repeats	Confirmation repeats
		Praise-word repeats
		Repair repeats
4	Key-word repeats	Lesson key-word repeats
		Activity key-word repeats
		Story key-word repeats
5	Approach-related repeats	Language-drill repeats
		Scaffolding repeats
6	Procedural repeats	Instruction-clarification repeats
		Classroom management repeats

With more varied data gleaned from private and international schools in the UAE and possibly with more classroom recordings from other countries where English is taught as a second language, it is more likely that the taxonomy would comprise more types and categories. Nevertheless, I am confident that this proposed taxonomy will provide a useful frame of reference for other researchers who are interested in investigating teacher talk and interaction in the L2 classroom. Indeed, with its different categories and types of repeats, the taxonomy

outlines an important characteristic of teacher talk that can help understand the complex relationship between interaction in the classroom and the pedagogical focus of the teacher.

6.3.4 A classroom-based English language proficiency assessment for teachers (CBELP)

As discussed in sub-section 6.2.1 of this chapter, a new assessment (CBELP) is suggested for use along with the IELTS test in order to establish the predictive validity of IELTS test scores for the graduation of the Bachelor of education students. This contribution has more than one facet. First, CBELP is a contribution to the field of language assessment as it offers a new format of assessing teachers' English language proficiency in the L2 classroom. This new format combines the IELTS test which measures general and academic language skills that are necessary for teachers to communicate and to function in professional contexts, and CBELP which assesses classroom-based language skills that are necessary for effective teaching in the L2 classroom. Second, the basis on which CBELP is introduced in this study contributes to the debate on language proficiency, whether it is one construct that can be assessed in the same way for all the domains of language use, or it is a variable construct that changes with the context, therefore it should be assessed in different ways. Third, this assessment contributes to the field of teacher education by drawing attention to the requirements of classroom-based language proficiency through the suggestion of a detailed rubric that takes into consideration the main areas of teacher talk and the characteristics of effective language use in the L2 classroom. In fact, the suggested way of assessing graduating B.Ed. students' English language proficiency can entice practitioners and policy makers to rethink their understanding of language proficiency as a general construct that functions in the same way in all contexts. It will, then, encourage them to adopt more flexible ways of assessing it for student teachers. In fact, this research thesis and the research study behind it can represent a resource for policy makers and for professionals to gain clear understanding of classroom-based language proficiency. It can provide them not only with evidence and explanations but also with alternatives to improve existing practice. The suggested CBELP assessment format and the available corpus of classroom teaching can help to establish a professional development program that promotes CBELP trainers and assessors. The corpus can provide trainees with ample opportunities to examine teacher talk in the classroom and to learn how to apply the suggested CBELP assessment in a classroom context.

6.3.5 A corpus of classroom discourse and the ‘mock’ IELTS speaking test for student teachers in the UAE

Another significant contribution of this research to the field of linguistics is the provision of a spoken corpus supported by transcripts of about 100,000 words (100,777) of 27 Emirati student teachers’ interviews for the ‘mock’ IELTS speaking test (about 40,000 words) and their performance as teachers in the classroom at different Emirati schools for young learners (about 60,000 words). The value of this corpus lies in being a spoken and a specialised one. By reviewing the literature on corpus linguistics, it was revealed that spoken corpora are not as available as written corpora due to different challenges in transcribing, especially with regard to pronunciation and non-verbal language (Kennedy, 1998; O’Keeffe & McCarthy, 2010; McEnery & Hardie 2012). Though in the case of the present study, transcription was done with no focus on such spoken features but just on lexical items regardless of the way they were produced, the gathered recordings remain available for further use for studies on spoken features of discourse on the test and in the classroom.

The corpus of this study is also specialised as it is collected for the specific purpose of the present research study (Kennedy, 1998), which is to analyse and compare student teachers’ language proficiency in the IELTS speaking test, and their performance in the classroom in order to evaluate the predictive validity of the test for a teacher education program. With its two main parts (the interview and the classroom teaching performance), the corpus offers valuable opportunities for practitioners in applied linguistics -including myself- to conduct further research into the L2 teacher’s lexical knowledge and use. This can encompass word formation, phraseology, collocations, idiomatic expressions, polysemous words, and so on (Moon, 2010). It can also be used to explore a plethora of areas covering teacher education and training like teaching and learning strategies, planning, questioning, giving feedback, reflective practice, and so on. The corpus can be useful as well to investigate controversies related to the L2 classroom in the specific context of the UAE and/or the Middle East in general. Some of those controversial issues are using the mother tongue versus the target language in the classroom, using formulaic language versus informal language, using language drills versus communicative activities, and so on. Finally, a corpus like this one can be used to inform L2 testing (Barker, 2010).

6.4 Limitations of the research study

Like any research study, despite the significance of the findings and the contribution to the existing body of knowledge, this research study has some limitations that could have played a negative role in misinforming some of the findings and probably misleading some conclusions. Three main limitations have been identified and reported to have had some potential impact on the findings of the present study: (1) having a small population size, (2) using a mock version of IELTS speaking test, and (3) using published materials of the IELTS speaking test.

First, the sample size in this study is 27 B.Ed. students from one college (Fujairah Women's college) which is a small population size compared with the number of B.Ed. students across the HCT system that includes 5 more colleges running the B.Ed. program. It is worth noting here that this limited number of subjects was rather caused by a cultural constraint than by any other logistic causes. As a matter of fact, the conservative background which most of the female student teachers come from, especially in the East Coast of the UAE where the college is located, does not allow them to share photos, videos, or audio-recordings with people outside their family circle even if it were for professional or academic purposes as it is the case in this research. Initially, I had planned to videotape classes of about 57 B.Ed. students which was the total number of the years 2, 3 and 4 students, but due to parents' and guardians' objections, I managed to get only 27 signed consent forms and only for audio-recordings.

This limitation can have some effect on the significance of the quantitative findings in terms of representativeness of the whole population. Therefore, my recommendation for future research is to try engage more participants from other colleges. However, I believe that having conducted a qualitative analysis using corpus linguistics and conversation analysis reduces the impact of this limitation on the final findings and conclusions for this research study.

Second, using a simulation of the IELTS speaking test by administering a mock copy of it could have had some negative impact on the subjects' performance as they were aware that it was not for official use. Indeed, some subjects scored clearly low in lexical diversity on the IELTS speaking test than on their classroom teaching performance (see table 4.1 for T03 and T14) which raises doubts that they were not taking the mock IELTS test seriously knowing that it had no importance for them. This could have also affected the quantitative analysis to some extent.

My recommendation would be to administer the mock test for student teachers who have not taken the official test yet, so they take it as a good practice for the real one.

The third possible limitation in this study is related to the use of published material to conduct the IELTS speaking test. This is a non-avoidable situation since there is no other way to get test materials that are not published to be sure that the subjects have not come across them before or even practised them. However, by asking the subjects -on an informal basis- if they have seen the test questions before, all of them answered that it was the first time for them to answer such questions. In fact, many of them complained that they could not say much because the main topics of test 1 and 4 (appendices 4 & 7) were about sports events and superstition, which they found difficult to talk about.

Finally, some classes were recorded by the students themselves by using software on their laptops, because as a researcher I found difficulty scheduling their classes at different times so I could attend and record everyone by myself. That was because they had a limited number of classes that they could teach and those classes were scheduled in coordination with their MCTs, which sometimes did not agree with my own schedule of visits to schools. The result is that 4 subjects did not realize that the recorder stopped in the middle of the class, so they had shorter recordings than the other participants. The usual range of time for classes is between 30 minutes for lower primary (grades 1-5) up to 50 minutes for upper grades (grade 6-9). However, for the subjects mentioned above their recordings were of the range 11 to 14 minutes (T11, T13, & T17) and 21 minutes for T12. My recommendation would be to assign the recordings to colleagues (MCTs) if it is impossible to schedule for all the classes to be recorded by the researcher.

6.5 Conclusion

This research study started from an argument calling into question the predictive validity of IELTS scores as a fundamental graduation requirement for the B.Ed. students at HCT, UAE. Questions were raised regarding the correspondence between the IELTS test task and the classroom teaching task to take the test scores as indicators of the classroom performance. Lexical diversity was used as an indicator of language proficiency to gauge the correlation

between the subjects' performances in both contexts. The obtained statistics revealed non-significant relationships between lexical diversity scores and between the IELTS speaking test scores and the subjects' grades on TP. The qualitative analysis demonstrated that the classroom context is clearly different than the test context which explains the discrepancies found in the quantitative analysis. Up to 16 types of teacher repeats in the classroom were identified that had direct impact on the lexical diversity of the subjects. As a result, a two-step validation of the IELTS scores is suggested in this study. First, a complementary classroom-based English language proficiency assessment (CBELP) is suggested to bridge the gap between the test task and the classroom teaching task. Second, a validation argument based on which IELTS and CBELP scores can be interpreted and high stake decisions can be made is developed.

The ultimate aim of the study was not to devalue the IELTS test, rather it was to raise awareness of the wrong use of its scores especially for high stake decisions as in the case of the B.Ed. students at HCT, UAE. Therefore, the research-based CBELP assessment is suggested as a means to strengthen the validity of IELTS scores as predictors of post-graduation language proficiency in the classroom. Its implementation is a way to improve the validity of the UAE B.Ed. qualification.

References

- Albrechtsen, D., Hasstrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Process and development*. Basingstoke: Plagrove Macmillan.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.
- Atkinson, J. M., & Heritage, J. (1984). Preference organization. In J. M. Atkinson & J. Heritage (Eds.). *Structures of social action: Studies in conversation analysis*. (pp.53-56). Cambridge: Cambridge University Press.
- Antaki, C. (2011). *An introduction to conversation analysis*. Retrieved July 7, 2015, from <http://homepages.lboro.ac.uk/~ssca1/notation.htm>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in practice*. Oxford: Oxford University Press.
- Barker, F. (2010). How can corpora be used in language testing. In A. O'Keefe & M. McCarthy. (Eds.). *The Routledge handbook of corpus linguistics*. (pp. 633-645). Abingdon: Routledge.
- Bell, A. (1984). Language style as audience design. *Language in society*, 13 (2), 145-204.

- Bialystok, E. (2001). *Bilingualism in development*. New York: Cambridge University Press.
- Bruner, J. (1983). *Child's talk*. Oxford: Oxford University Press.
- Canale, M. & Swain, M. (1981). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47
- Cazden, C. B. (1988). *Classroom discourse*. Portsmouth: Heinemann.
- Chamberland, C., Saint-Aubin, J. & Légère, M. (2013). The impact of text repetition on content and function words during reading: Further evidence from eye movements. *Canadian Journal of Experimental Psychology*, 67(2), 94-99.
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson. (Eds.). *The Routledge handbook of language testing*. (pp.21-33). Oxon: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3-13.
- Chaudron, C. (1988). *Second language classroom*. Cambridge: Cambridge University Press.
- Cook, G. (1994). Repetition and learning by heart: an aspect of intimate discourse, and its implications. *ELT Journal*, 48(2), 133-141.
- Cotton, F. & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a group of international students studying at the University of Tasmania. *IELTS Research Reports Volume 1*. 72-115.

- Coulthard, M. (1985). *An introduction to discourse analysis*. (2nd ed.). Harlow: Pearson Education Limited.
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. Thousand Oaks: Sage.
- Creswell, J. W. & Clark, V. P. (2011). *Designing and conducting mixed methods research*. (2nd ed.). Thousand Oaks: Sage.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.). *Educational Measurement*. (2nd ed.). (pp. 443-507). Washington: American Council on Education.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*. 52(4), 281-302.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243-263.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561-580.
- Cummins, J. (2000). *Language, power and pedagogy*. Clevedon: Multilingual Matters.
- Daller, H., Van Hout, R. & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics*, 24(2), 197-222.
- Daller, H., & Xue, H. (2007). Lexical richness and oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.). *Modelling and*

- assessing vocabulary knowledge*. (pp.150-164). Cambridge: Cambridge University Press.
- Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.
- Davies, A. (2008). *Studies in language testing 23: Assessing academic English*. Cambridge: Cambridge University Press.
- Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.). *Handbook of research on teaching*. (3rd ed.). New York: Macmillan.
- Duff, P. (2000). Repetition in foreign language classroom interaction. In J. K. Hall and L. S. Verplaetse (Eds.). *The Development of second and foreign language learning through classroom interaction*. (pp. 109–38). Lawrence Erlbaum.
- Duranti, A. (1992). Language in context and language as context: The Samoan respect vocabulary. In C. Goodwin, & A. Duranti (Eds.). *Rethinking context*. (pp.77-99) Cambridge: Cambridge University Press.
- Duran, P., Malvern, D., Richards, B. & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, 25(2), 220-242.
- Elder, C. (1993). Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing*, 2(1), 68-89.
- Ellis, R. (2008). *The study of second language acquisition*. (2nd ed.). Oxford: Oxford University Press.

- Field, J. (2011). Cognitive validity. In L. Taylor (Ed.). *Studies in language testing 30: Examining speaking*. (pp. 65-111) Cambridge: Cambridge University Press.
- Freeman, D., Katz, A., Gomez, P. G., & Burns, A. (2015). English-for-Teaching: Rethinking teacher proficiency in the classroom. *ELT Journal*, 69(2), 129-139.
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Education.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: an advanced resource book*. Oxon: Routledge.
- Galaczi, E. & Ffrench, A. (2011). Context validity. In L. Taylor (Ed.). *Studies in language testing 30: Examining speaking*. (pp. 112-170) Cambridge: Cambridge University Press.
- Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.
- Gass, S. M., Mackey, A. & Pica, T. (1998). The role of input and interaction in second language acquisition: introduction to the special issue. *Modern Language Journal*, 82(3), 299-307.
- Ghasemi, A. & Zahediasl, S. (2012). Normality Tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology Metabolism*, 10(2), 486-489.
- Goodwin, C., & Duranti, A. (1992). Rethinking context: An introduction. In C. Goodwin, & A. Duranti (Eds). *Rethinking context*. (pp.1-42). Cambridge: Cambridge University Press.
- Halliday, M. A. K. & Hassan, R. (1976). *Cohesion in English*. London: Longman.

- Hatch, E. (1978). Discourse analysis and second language acquisition. In E. Hatch (Ed.). *Second language acquisition*. Rowley, MA: Newbury House.
- Hellerman, J. (2003). The interactive work of prosody in the IRF exchange: teacher repetition in feedback moves. *Language in Society*, 32, 79-104.
- Heritage, J. (1997). Conversation analysis and institutional talk: Analysing data. In D. Silverman (Ed.). *Qualitative research: Theory, method and practice*. (pp.161-182). London: Sage.
- Horst, M. (2009). Revisiting classrooms as lexical environments. In T. Fitzpatrick & A. Barfield (Eds.). *Lexical processing in second language learners*. Retrieved from <https://www.dawsonera.com/abstract/9781847691538>
- Hutchby, I., & Wooffitt, R. (1998). *Conversation analysis*. Cambridge: Polity Press.
- IELTS. (2016). *IELTS*. Retrieved on August 15, 2016 from <http://www.ielts.org> .
- IELTS. (2007). *Handbook 2007*. Cambridge: University of Cambridge ESOL Examination.
- IELTS. (2007). *Official IELTS practice materials*. Cambridge: University of Cambridge ESOL Examination.
- Ingram, D. & Bayliss, A. (2007). IELTS as a predictor of academic language performance, part1. *IELTS Research Reports Volume 7*, 137-204.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84.

- Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, 63(1), 87-106.
- Kane, M. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson. (Eds.). *The Routledge handbook of language testing*. (pp.21-33). Oxon: Routledge.
- Kane, T. M. (1992). An argument-based approach to validity. *Psychological Bulletin*. 112(3), 527-535.
- Kane, T. M. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342.
- Kane, T. M. (2011). Validating score interpretation and uses: Messick lecture, language testing research colloquium, Cambridge April 2010. *Language Testing*, 29(1), 3- 17.
- Kao, S. M., Carkin, G., & Hsu, L. F. (2011). Questioning techniques for promoting language learning with students of limited L2 oral proficiency in a drama oriented language classroom. *The Journal of Applied Theatre and Performance*, 6(4), 489-515.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Harlow, Essex: Longman.
- Kerstjens, M. & Nery, C. (2000). Predictive validity in the IELTS test: a study of the relationship between IELTS scores and students' subsequent academic performance. *IELTS Research Reports Volume 3*, 85-108
- Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.

- Lightbown, P. M. & Spada, N. (2006). *How languages are learned*. (3rd ed.). Oxford: Oxford University Press.
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126-141.
- Long, M. & Sato, C. (1983). Classroom foreign talk discourse: Forms and functions of teachers' questions. In H. W. Seliger & M. H. Long (Eds.). *Classroom-oriented research in second language acquisition*. (pp. 268-85). Rowley: Newbury House.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- MacWhinney, B. (2010). *The CHILDES project*. Retrieved on October 15, 2010 from <http://childes.psy.cmu.edu/manuals/CHAT.pdf>
- Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan, & A. Wray (Eds.). *Evolving models of language*. (pp.58-71). Clevedon: Multilingual Matters.
- Malvern, D. D., & Richards, B. J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85-104.
- Malvern, D., Richards, B. J., Chipere, N., & Dúran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Malvern, D. D., & Richards, B. J. (2009). A new method of measuring rare word diversity: The example of L2 learners of French. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.). *Vocabulary studies in first and second language acquisition*. (pp.164-178). Hampshire: Palgrave Macmillan.

- McCarthy, M. & O’Keefe, A. (2010). What are corpora and how have they evolved? In A. O’Keefe & M. McCarthy. (Eds.). *The Routledge handbook of corpus linguistics*. (pp. 3-13). Abingdon: Routledge.
- McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing*, 24(4), 459-488.
- McCarthy, P. M. & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behaviour Research Methods*, 42(2), 381-392.
- McEneaney, T. & Hardie, A. (2012). *Corpus linguistics*. Cambridge: Cambridge University Press.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323-337.
- McLaughlin, B. (1985). *Second language acquisition in childhood: Volume 2 school age*. (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
- Meara, P. M., Lightbown, P. M. & Halter, R. H. (1997). Classrooms as lexical environments. *Language Teaching Research*. 1(1), 28-47.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. USA: Harvard University Press.
- Messick, S. (1989). Validity. In R. L. Lynn (Ed.). *Educational measurement*. (3rd ed.). (pp. 13-103). New York: American Council on Education and Macmillan.

- Moon, R. (2010). What can a corpus tell us about lexis? In A. O’Keefe & M. McCarthy. (Eds.). *The Routledge handbook of corpus linguistics*. (pp. 197-211). Abingdon: Routledge.
- Myhill, D., Jones, S., & Hopper, R. (2006). *Talking, listening, learning: Effective talk in the primary classroom*. Berkshire: Open university Press.
- O’Connor, M.C. and Michaels. S. (1993). Aligning academic task and participation status through revoicing: analysis of a classroom discourse strategy. *Anthropology and Education Quarterly*, 24(4), 318-35.
- Oller, J. W. Jr. (1979). *Language tests at school*. London: Longman.
- Oller, J. W. Jr. (1980). Language testing research (1979-1980). *Annual Review of Linguistics*, 1(1), 124-150.
- Park, Y. (2014). The roles of third-turn repeats in two L2 classroom interactional contexts. *Applied Linguistics*, 35(2), 145-167.
- Paul, A. (2007). IELTS as a predictor of academic language performance, Part 2. *IELTS Research Reports* 7, 205-271.
- Peregoy, S. F. & Boyle, O. F. (2005). *Reading, writing and learning in ESL*. (4th ed.). Boston, MA: Pearson Education.
- Pica, T. (1994). Research on Negotiation: What does it reveal about second language learning conditions, processes, and outcomes? *Language Learning*, 44(3), 493-527.
- Piirainen-Marsh, A. and Tainio, L. (2009). Other-repetition as a resource for participation

- in the activity of playing a video game. *The Modern Language Journal*, 93(2), 153-69.
- Rea-Dickins, P., Kiely, R., & Yu, G. (2011). Uses and impact of test scores in university admissions processes: The language test as the 'hard' criterion. In B. O'Sullivan (Ed.). *Language testing: Theories and practices*. (pp.262-281). Hampshire: Palgrave Macmillan.
- Read, J., & Nation, P. (2006). An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Reports* 6, 207-231.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Rumsey, L. (2013). *Shedding light on the predictive validity of English proficiency tests in predicting academic success* (Doctoral dissertation). Retrieved from ETHOS database (<http://ethos.bl.uk/SearchResults.do>).
- Rydland, V. & Aukrust, V. G. (2005). Lexical repetition in second language learners peer play interaction. *Language Learning*, 55(2), 229-74.
- Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Schegloff, E. A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26(1), 99-128.
- Schegloff, E., A., Jefferson, G., & Sacks, H. (1977). Preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361-382.
- Schmitt, N. (2010). *Researching vocabulary*. Hampshire: Palgrave Macmillan.
- Scrivener, J. (2011). *Learning teaching*. (3rd Ed.). London: Macmillan.

- Scott, M. (2010). *WordSmith tools*. (Version 5.0). Retrieved from <http://www.lexically.net/wordsmith/version5/index.html>
- Seedhouse, P. (2004). The interactional architecture of the language classroom: a conversation analysis perspective. *Language Learning*, 54(S1), 1-262.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education*, 19, 405-450.
- Sinclair, J. M. & Brazil, D. (1982). *Teacher talk*. Oxford: Oxford University Press.
- Sinclair, J. M. & Coulthard, R. M. (1975). *Towards an analysis of discourse*. Oxford: Oxford University Press.
- Taylor, L. (2001a). Revising the IELTS speaking test. *Research Notes*, 4, 9-11.
- Taylor, L. (2001b). Revising the IELTS speaking test: developments in test format and task design. *Research Notes*, 5, 2-5.
- Taylor, L. (2001c). Revising the IELTS speaking test: retraining IELTS examiners worldwide. *Research Notes*, 6, 9-11.
- Taylor, L. (2003). The Cambridge approach to speaking assessment. *Research Notes*, 13, 2-4.
- Tashakkori, A. & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks: Sage.
- Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oak: Sage.
- ten Have, P. (1999). *Doing conversation analysis: A practical guide*. London: Sage.

- Tidball, F., & Treffers-Daller, J. (2007). Exploring measures of vocabulary richness in semi-spontaneous French speech. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.). *Modelling and assessing vocabulary knowledge*. (pp.133-149). Cambridge: Cambridge University Press.
- Treffers-Daller, J. (2009). Language dominance and lexical diversity: how Bilinguals and L2 Learner differ in their knowledge and use of French lexical and functional items. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.). *Vocabulary studies in first and second language acquisition*. (pp.74- 90). Hampshire: Palgrave Macmillan.
- Tomlin, K. (1994). Repetition in second language acquisition. In B. Johnstone (Ed.). *Repetition in discourse: Interdisciplinary perspectives. vol. 1*.(pp. 172-94). Ablex.
- Tsui, A. (1992). A functional description of questions. In M. Coulthard (Ed.). *Advances in spoken discourse analysis*. London: Routledge.
- van Lier, L. (1988). What's wrong with classroom talk? *Prospects*, 3(3), 267-283.
- van Lier, L. (1996). *Interaction in the language classroom*. Harlow: Pearson Education.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17(1), 65-83.
- Walsh, S. (2011). *Exploring classroom discourse*. Oxon: Routledge.
- Walsh, S. (2013). *Classroom discourse and teacher development*. Edinburgh: Edinburgh Press.

- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*.
Basingstoke: Palgrave Macmillan.
- White, J. & Lightbown, P. M. (1984). Asking and answering in ESL classes. *Canadian Modern Language Review*, 40(2), 228-244.
- Yap, B. W. & Sim, C. H. (2011). Comparison of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12), 2141–2155.
- Yifat, R. & Zardunaisky-Ehrlich. (2008). Teachers' talk in preschool during circle time: the case of revoicing. *Journal of Research in Childhood Education*, 23(2), 211-226.
- Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics*, 1-24.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency. *Studies in Second Language Acquisition*, 27(4), 567-595.