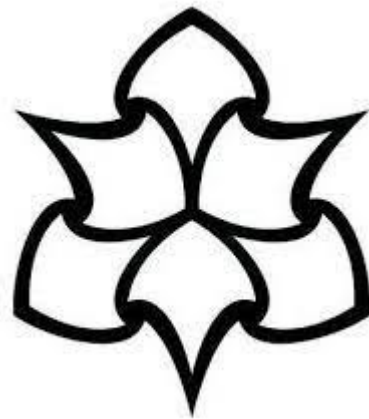


The Design and Implementation of a Framework for the Analysis of Lexical Errors in the Compositions of Students of English as a Second Language

ANTHONY JOHN PICOT

A thesis submitted in partial fulfilment of the requirements of
the
Manchester Metropolitan University for the Degree of Doctor
of Philosophy

Department of Languages, Information and Communications
Manchester Metropolitan University



2023

Declaration

I confirm that no material contained in the thesis has been used in any other submission for another academic award.

Abstract

This dissertation seeks to revisit the potential of Error Analysis with a specific focus on lexical errors which will inform second language teaching and learning. The dissertation attempts to devise a new framework and guidance for analysing lexical errors, as a systematic and reliable method of identifying and categorising lexical errors would benefit SLA researchers, English Language teachers and learners (James, 1998; Hemchua and Schmitt, 2006). A replication of one of the most recent and potentially strongest Lexical Error Analysis (LEA) studies, Hemchua and Schmitt (2006), was completed and similarity was found between sets of results, giving promise to the concept and replicability of LEA. Other previous frameworks for LEA, namely Dušková (1969), Corder (1973), Richards (1971), Zimmerman (1986), Zimmerman (1987), Meara and English (1987), Lennon (1991), Zughoul (1991), Engber (1995), James (1998), Hemchua and Schmitt (2006) and Llach (2011), were tested and analysed, for their strengths and weaknesses in terms of their ease of use and the depth of analysis that they provide, before a new LEA framework was devised (NewLEAF Version 1). The new framework was then tested: interrater agreement for error identification and categorisation was investigated in two studies. The first, which showed promising quantitative results for consistent identification and categorisation between raters, also sought qualitative feedback on using the framework and accompanying guidance when it was employed by six highly qualified and highly-experienced EAP university teachers. Following this feedback, refinements were made to the framework and guidance to produce NewLEAF2, which was then tested on 41 participants. These participants comprised a mixture of Linguistics and TESOL staff and students. Following less encouraging results in terms of error identification and categorisation, further refinements were made to produce NewLEAF3, which was then used to analyse 20 scripts produced by Greek learners of English. It was found that the framework and guidance proved to be easy to use in that there were no issues in error identification, no uncategorisable errors, no dual categorisation issues, and a satisfactory spectrum of lexical error types. Results of the final analyses showed that there were far more semantic errors than formal errors: phrase errors were the most common followed by

preposition and verb errors. Although further work may be required to improve the framework in terms of potential issues with inter-rater agreement in the areas of error identification and categorisation, the improved depth of analysis and ease of use provided by the new framework offers considerable advantage to learners and teachers of English.

Table of contents

Declaration	2
Abstract	2
Table of contents	4
Table of Tables	13
List of abbreviations	15
Acknowledgments	15
Chapter 1 - Introduction	16
1.1 Aims of the dissertation.....	16
1.2 Background and justification for study	16
1.2.1 Uses, users and contexts for the LEA tool	18
1.2.1.1 Learners	18
1.2.1.2 Teachers.....	19
1.2.1.3 Markers.....	19
1.2.1.4 Materials writers	20
1.2.1.5 Second Language Researchers	20
1.3 Organisation of the dissertation	20
Chapter Two: Literature Review	23
2.0 Introduction	23
2.1. Definition of errors in language learning and teaching.....	25
2.1.1 Justification for error correction	26
2.2 Error analysis introduction	28
2.2.1 Definition and purposes.....	28
2.2.2 The need for EA	29
2.2.2.1 Academic/ SLA research.....	30
2.2.2.2 Pedagogic uses.....	30
2.2.3 Other uses of EA	31
2.2.4 Computer-aided EA	32
2.3 A history of EA	34

2.3.2 The rise of EA.....	35
2.4 The process of EA	36
2.4.1 Collection of errors.....	36
2.4.2 Identification of errors	37
2.4.3 Description of errors	41
2.4.3.1 Frameworks for error categorisation	45
2.4.4 Quantification of errors	49
2.4.5 Explanation of errors.....	50
2.4.6 Evaluation of errors.....	53
2.4.7 Summary of the limitations of EA.....	55
2.5. Lexical error analysis	60
2.5.1 The importance of lexis.....	60
2.5.2 Lexis and grammar	61
2.5.3 Lexicogrammar	62
2.5.4 Lexical error analysis vs whole language error analysis	63
2.5.5 A justification for LEA.....	63
2.5.6 Definition of a lexical error	66
2.5.7 Taxonomies in the published LEA frameworks	69
2.5.7.1 Dušková (1969) Etiologic/process-oriented.....	71
2.5.7.2 Richards (1971) Origin of influence criterion	73
2.5.7.3 Corder (1973) Grammatical or linguistic criterion	76
2.5.7.4 Zimmerman (1986) Descriptive/product-oriented	77
2.5.7.5 Zimmerman (1987) Form-and content-oriented/product-oriented ...	80
2.5.7.6 Meara and English's Framework (1987).....	82
2.5.7.7 Lennon (1991) Grammatical or linguistic criterion/word-class criterion/product-oriented	83
2.5.7.8 Zughoul (1991) Etiologic/product/process-oriented.....	84
2.5.7.9 Engber (1995) Etiologic/product-oriented.....	86
2.5.7.10 James (1998) Form and content-oriented	87
2.5.7.11 Hemchua and Schmitt (2006) Form and content oriented.....	89
2.5.7.12 Llach (2011) Form/content/origin of influence criterion	92
2.6 Summary/conclusion.....	94

Chapter 3 (Study 1) Replication.....	96
3.1 Introduction	96
3.2 Rationale for study.....	97
3.3 Methodology	99
3.3.1 Research questions.....	99
3.3.2 Participants.....	99
3.3.3 Ethics.....	99
3.3.4 Analysis	100
3.4 Results.....	101
3.4.1 Word count and standard deviation	102
3.4.2 Error count.....	102
3.4.3 Types of errors made.....	103
3.4.3.1 Formal and semantic errors	103
3.4.3.2 Most common errors overall.....	107
3.4.4 Summary of results.....	109
3.5 Discussion of issues in using the 2006 framework	110
3.5.1 Problems of error identification	110
3.5.2 Grammatical vs lexical error	110
3.5.3 Problems of LE classification.....	110
3.5.3.2 Semantic errors.....	112
3.5.3.3 Error Count and Resulting Analysis Issues	115
3.5.3.4 Miscellaneous Issues	116
3.6 Implication for practice	117
3.7 Conclusion	117
Chapter 4 (Study 2) Learning from Existing Lexical Error Analysis	
Frameworks: A Comparative Approach.....	120
4.1 Introduction	120
4.2 Methodology	121
4.2.1 Research questions.....	122
4.2.2 Participants.....	122

4.2.3 Ethics.....	123
4.2.4 Analysis	123
4.2.4.2 Process	123
4.3 Results and discussion	124
4.3.1 Dušková (1969) (Etiologic/process-oriented).....	124
4.3.1.1 Discussion of results	125
4.3.1.2 Ease of use	125
4.3.1.3 Depth of analysis.....	125
4.3.2 Corder 1973 (Grammatical or linguistic criterion)	127
4.3.2.1 Discussion of results	127
4.3.2.2 Ease of use	127
4.3.2.3 Depth of analysis.....	128
4.3.2.4 Points to consider when creating a new framework for LEA	129
4.3.3 Zimmerman (1986) (Descriptive/product-oriented).....	129
4.3.3.1 Discussion of results	129
4.3.3.2 Ease of use	130
4.3.3.3 Depth of analysis.....	131
4.3.3.4 Points to consider when creating a new framework for LEA	131
4.3.4 Engber (1995) (Etiologic/product-oriented).....	131
4.3.4.1 Discussion of results	132
4.3.4.2 Ease of use	133
4.3.4.3 Depth of analysis.....	133
4.3.4.4 Points to consider when creating a new framework for LEA	134
4.3.5 James (1998) (form- and content-oriented)	134
4.3.5.1 Discussion of results	136
4.3.5.2 Ease of use	136
4.3.5.3 Depth of analysis.....	137
4.3.5.4 Points to consider when creating a new framework for LEA	138
4.3.6 Hemchua and Schmitt (2006).....	138
4.3.6.1 Discussion of results	139
4.3.6.2 Ease of use	139
4.3.6.3 Depth of analysis.....	140
4.3.6.4 Points to consider when creating a new framework for LEA	140
4.3.7 Richards (1971) (Origin of influence or cause criterion)	141

4.3.8 Meara and English (1987)	141
4.3.9 Zughoul (1991)	141
4.3.10 Lennon (1991)	142
4.3.11 Llach (2011).....	142
4.4 Comparison of results	142
4.5 Conclusions	144
Chapter Five (Study 3) Description and testing of a New Lexical Error	
Analysis Framework (NewLEAF1).....	145
5.0 Introduction	145
5.1 Description of NewLEAF1	146
5.1.2 The categories.....	150
5.1.2.1 Section A.....	150
5.1.2.1 Section B.....	150
5.2 Variation - whole language analysis, including grammar, phonology and punctuation.....	154
5.3 Adaptability	154
5.4 Analysis	155
5.4.1 Guidance for analysts using the new LEA framework.....	155
5.5 Results.....	156
5.5.1 Results of the lexical error analysis	156
5.5.2 Comparison of performance with other frameworks	158
5.6 Summary/conclusions.....	159
5.7 Limitations of the new framework.....	159
Chapter 6 (Study 4) Testing and refining NewLEAF1	161
6.1 Introduction	161
6.1.1 Aims.....	161
6.1.2 Research questions.....	162
6.2 Methodology	162
6.2.1 Methodological Approach	163

6.2.2 Ethical considerations.....	163
6.2.3 Participants.....	164
6.2.3.1 Sampling.....	164
6.2.3.2 Students.....	164
6.2.3.3 Teachers.....	164
6.2.4 Data collection.....	165
6.2.4.1 Collection of the scripts.....	165
6.2.4.2 Lexical error analysis of the script.....	165
6.2.4.3 Think-aloud protocol.....	167
6.2.4.4 Observation of participants.....	169
6.2.4.5 Semi-structured individual interviews.....	169
6.2.5 Data analysis.....	170
6.2.5.1 Research question 1.....	171
6.2.5.2 Research question 2.....	173
6.2.5.3 Research questions 3 and 4.....	173
6.2.5.4 Research question 5.....	173
6.3 Results and discussion.....	175
6.3.1 Research question 1.....	175
6.3.2 Research question 2.....	184
6.3.3 Research question 3.....	189
6.3.4 Research question 4.....	190
6.3.5 Research question 5.....	192
6.3.5.1 Analysis of results and recordings of identification and categorisation tasks.....	192
6.3.5.2 Analysis of semi-structured interview data.....	192
6.4 Summary of recommendations for refinement, based on the results and discussion.....	194
6.4.1 Suggestions to be adopted.....	195
6.4.1.1 Clarify terminology in the framework and guidance.....	195
6.4.1.2 Remove ambiguity in framework and guidance.....	195
6.4.1.3 Reduce potential for dual categorisation.....	196
6.4.1.4 Provide clearer advice for categorisation in the guidance.....	197
6.4.2 Suggestions not adopted.....	198

6.4.2.1 Remove ambiguity in framework and guidance.....	198
6.4.3 Other observations	199
6.5 Conclusion	200
6.5.1 Summary	200
6.5.2 Limitations of current study.....	201
6.5.2.1 Methodological limitations	201
6.5.2.2 Issue with the framework.....	202
6.5.3 The next steps	202
Chapter 7 Study 5: Testing and Refining NewLEAF2	203
7.1 Introduction	203
7.2 Methodology	204
7.2.1 Research questions.....	204
7.2.2 Part One	205
7.2.2.1 Sampling, participants and ethics.....	205
7.2.2.2 Data collection.....	205
7.2.3 Part 2	207
7.2.3.1 Sampling	207
7.2.3.2 Participants	207
7.2.3.3 Ethics	210
7.2.3.4 Data collection.....	210
7.2.3.5 Data analysis.....	211
7.3 Results and discussion	213
7.3.1 Part 1	213
7.3.1.1 Identification research question 1.....	213
7.3.1.2 Categorisation Research question 2	214
7.1.3.3 Research aim.....	216
7.3.2 Part Two	216
7.3.2.1 Identification research question 1.....	216
7.3.2.2 Categorisation research question 2.....	220
7.3.2.3 Improvements to the framework research aim	222
7.4 Conclusion	233
7.4.1 Identification	233

7.4.2	Categorisation	235
7.4.3	Limitations of Study 5	236
Chapter 8 Study 6 Implementation of NewLEAF3		239
8.1	Introduction	239
8.2	Research questions	239
8.3	Methodology	239
8.4	Results and discussion	240
8.5	Conclusions	247
Chapter 9 Conclusion.....		250
9.1	Summary and conclusions.....	250
9.1.1	Literature review	250
9.1.2	Study 1	251
9.1.3	Study 2	252
9.1.4	Study 3	253
9.1.5	Study 4	253
9.1.6	Study 5	254
9.1.6.1	Identification	254
9.1.6.2	Categorisation	256
9.1.7	Study 6	257
9.2	Description and limitations of NewLEAF3	258
9.2.1	Description.....	258
9.2.2	Limitations	259
9.2.2.1	Identification	259
9.2.2.2	Categorisation	259
9.2.2.3	General	259
9.3	Benefits of NewLEAF3.....	260
9.4	Future improvements to NewLEAF3	261
9.5	Future hopes for NEWLEAF and LEA research.....	261
9.6	Final thoughts	264
References.....		265

Appendices.....	278
Appendix 3.1 - Consent form	278
Appendix 5.1 Instructions for analysts (NewLEAF1).....	279
Appendix 6.1 The new guidance and framework (NewLEAF2) Lexical error analysis - instructions for analysts.....	283
Appendix 6.2 Letter of ethical approval.....	289
Appendix 6.3 Participant information sheet for teachers 27/12/18 Version 2290	
Appendix 6.4 Consent form for students Version 1.....	292
Appendix 6.5 Consent form for teachers Version 3	293
Appendix 6.6 Student's essay.....	294
Appendix 6.7 - Lexical error analysis - instructions for participants	295
Appendix 6.8 Interview schedule for participants using the new framework	308
Appendix 6.9.1 P1P2 LEA Transcript	309
Appendix 6.9.2 P3P4 LEA transcript.....	318
Appendix 6.9.3 P5P6 LEA transcript.....	335
Appendix 6.9.4 Post LEA interview P2	343
Appendix 6.9.5 Post LEA interview P3	348
Appendix 6.9.6 Post LEA interview P5	353
Appendix 7.1 Participant information sheet.....	357
Appendix 7.2 Lexical error analysis of international students' compositions consent form	359
Appendix 7.3 Printed or emailed instructions for data collection.....	360
Appendix 7.4 Letter of ethical approval.....	370
Appendix 7.5 Questionnaire.....	371
Appendix 7.6 Emails to participants.....	372
Appendix 7.7 Email 2	373
Appendix 7.8 Table 7.3 Part One Error Identification.....	374
Appendix 7.9 Table 7.4 Part One error categorisation and confidence scores	

Appendix 7.10 Table 7.5 Part One error categorisation with mode.....	379
Appendix 7.11 Table 7.6 Part Two Group A and Group B error identification.....	381
Appendix 7.12 Table 7.7 Part Two Group A error identification	383
Appendix 7.13 Table 7.8 Part Two Group B error identification	387
Appendix 7.14 Table 7.9 Part Two Group A and Group B error categorisations with mode.....	391
Appendix 7.15 Table 7.10 Part Two Group A and B categorisation average confidence scores	392
Appendix 7.16 Lexical error guidance and framework (NewLEAF3).....	394

Table of Tables

Table 1 Corder's (1973:278) Surface Error Taxonomy	45
Table 2 James' (1998:274) Framework for Error Analysis (target modification taxonomy).....	47
Table 3 Dušková's (1969:35) Framework for EA	72
Table 4 Distribution of LEs in Dušková (1969:35).....	72
Table 5 Richards' (1971:7-15) Framework for EA.....	74
Table 6 Richards' (1971:24-35) Framework for EA: Further sub-categorisation	75
Table 7 Corder's (1973:278) Framework for EA	76
Table 8 Zimmerman's (1986:32-36) Framework for LEA	79
Table 9 Zimmermann's (1987:58-62) Framework for LEA	81
Table 10 Meara and English's (1987:4) Framework for LEA.....	82
Table 11 Lennon's (1991:34-39) Framework for EA	83
Table 12 Zughoul's (1991:47) Framework for LEA	85
Table 13 Engber's (1995: 146) Framework for LEA.....	87
Table 14 James' (1998:274) Framework for EA.....	88
Table 15 Hemchua and Schmitt's (2006: 12) Framework for LEA	91
Table 16 Llach's (2011:123-124) Framework for LEA.....	92
Table 17 Hemchua and Schmitt's (2006:12) framework for LEA	101
Table 18 Word and error count in the two studies.....	102
Table 19 Summary of frequency in formal and semantic errors.....	103
Table 20 Rank-order frequency of LEs	106

Table 21 Number of errors in each category and the number of uncategorysable errors using Dušková (1969)	124
Table 22 Number of errors in each category and the number of uncategorysable errors using Corder (1973)	127
Table 23 Number of errors in each category and the number of uncategorysable errors using Zimmerman (1986)	129
Table 24 Number of errors in each category and the number of uncategorysable errors using Engber (1995).....	132
Table 25 Number of errors in each category and the number of uncategorysable errors using James (1998).....	136
Table 26 Number of errors in each category and the number of uncategorysable errors using Hemchua and Schmitt (2006).....	139
Table 27 Comparison of the number of errors identified by the various LEA frameworks and the numbers of categorisable vs uncategorysable errors	142
Table 28 NewLEAF1	149
Table 29 Results of LEA on Greek Data using NEWLEAF1	157
Table 30 Comparison of the number of categorisable vs uncategorysable errors identified by Version 1 and previously published frameworks	158
Table 31 Research questions, data sets used to answer questions and purpose	162
Table 32 Summary of stages of data analysis	166
Table 33 Total number of errors identified by different participants	179
Table 34 Instances of rater categorisation agreement by different participants	180
Table 35 Quantitative analysis of emergent codes for categorisation issues	184
Table 36 Themes extracted from the transcription that relate to ease of use	189
Table 37 Themes extracted from the transcription that relate to depth of analysis	190
Table 38 Themes extracted from the transcription that relate to how analysis findings could be used in their practice	192
Table 39 Themes extracted from the transcription that relate to how the guidance and framework could be improved	194
Table 40 Participants in Study 5	208
Table 41 Comparison of error identification with NewLEAF1	214
Table 42 Dual categorisation issues, codes and frequencies	224

Table 43 Results of LEA on Greek Data using NewLEAF3	240
Table 44 Comparison of lexical errors found using Version 1 and Hemchua and Schmitt's (2006) framework when using the Greek data	241
Table 45 Frequency of formal and semantic errors in Study 1 and Study 6242	
Table 46 Types and frequencies of the LE s made in Study 1 and Study 6245	

List of abbreviations

EA	Error Analysis
EFL	English as a Foreign Language
ESL	English as a Second Language
IELTS	International English Language Testing System
FL	Foreign Language
LEA	Lexical Error Analysis
LE	Lexical Errors
L1	The first language
L2	The second language/the foreign language
SLA	Second Language Acquisition

Acknowledgments

I would like to thank my supervisor, Dr. Huw Bell, for his patience and expertise, my line managers, Dr. Derek Bousfield and Mr. Richard Silburn, for their kind encouragement and support in completing this dissertation. I would also like to thank the participants for their help in testing out different versions of the framework, and finally, I would like to thank my family for their love.

Throughout the dissertation, formatting is used as follows:

Extracts of student writing that contain errors are presented in italics with an asterisk before the error: e.g. **My dog is smelling terrible.*

The categories used to classify the errors are presented in square brackets: e.g. [Prepositional Partner]

Corrected versions of errors are marked '<': e.g. **My dog is smelling terrible.*
<My dog smells terribly.

Chapter 1 - Introduction

Throughout my career as an English Language teacher, and my reading to support my professional development, I have become increasingly aware of how lexis is the organising principle in much of what we say, write, hear and read: more so than grammar, as many believe. And yet, grammar retains the organising role in coursebooks, and, in my experience, is still the preoccupation of many learners and teachers. This is unsurprising given the unwieldy lexicon of English, its size and various ways we have of classifying it and the rather more finite and straightforward grammatical system. Whenever I am asked to mark the written work of learners, without guidance, I find it quite subjective to decide whether a lexical error has been made and if so, what type of error it is. It is these issues that first attracted me to start investigating LEA and how it might be refined, improved upon or even standardised. It occurred to me that if this were possible, it may inform a number of areas of applied linguistics, such as marking in high stakes exams where lexical resource is a criterion, feedback to learners to inform them of their errors, data for teachers to inform remedial teaching, and also data for second language acquisition researchers that are specifically interested in the acquisition of lexis.

1.1 Aims of the dissertation

This dissertation aims to improve on existing Lexical Error Analysis (LEA) schemes and procedures by developing and testing a framework and accompanying guidance that can be used by both teachers and Second Language Acquisition (SLA) researchers to ascertain the type and frequency of lexical errors made in the written compositions of learners of English as a second or additional language. This information will be of use to learners, teachers and SLA researchers.

1.2 Background and justification for study

Growing numbers of immigrants are entering education systems around the world where they are studying in a second language. This represents a major challenge as these learners often require help developing academic literacy skills (Doolan and Miller, 2012). Therefore, it is important to find ways to bridge this gap as this will lead to better retention and learning outcomes. This can be done by identifying the errors that these groups make (*ibid* 2012). Learner error has always been of central importance to language teachers, and has been considered extensively by

applied and theoretical linguists since at least the 1940s. Error Analysis (EA), a method of analysing errors in language learner output, was first systematised by Pit Corder (1967), who identified its practical and theoretical value as a focus for remedial teaching and materials design. In my experience, learners themselves are also very interested in finding out the type and frequency of the errors that they make. EA has already contributed to SLA research in that it helps us to understand how SLA proceeds and where it does not, its stages and what processes are at stake, and where in terms of acquisition/learning the learner is (Llach, 2011: xi-xiii). However, a more reliable method of conducting LEA will further contribute as a tool in SLA studies. EA fell out of fashion towards the latter part of the 20th century due to a number of recurrent and apparently intractable difficulties and because of changes in the way that 'errors' are viewed. However, Taylor (1986) and Lennon (1991) have suggested ways that earlier difficulties with EA may be overcome. It has recently re-emerged in the form of computer-aided EA where large corpora of student compositions can be tagged for error either manually or by automated computer software (Gamallo *et al*, 2015) to analyse type and frequency of errors made by second language learners. However, current tagsets, or categories used to group lexical error types, do not appear to provide for a particularly wide spectrum of errors. Therefore, before a more comprehensive analysis of learner errors can be undertaken, a more detailed way of grouping them should be found. There has been some resurgence of interest in lexical errors in the last 20 years (e.g. Tschichold, 2003, Hemchua and Schmitt, 2006 and Llach 2005, 2007a, 2007b, 2011 Al-Shormani and Al-Sohbani, 2012, Shaalani, Magdy, and Fahmy, 2015, Lee, 2017, Picot, 2017). This re-emergence is part of a wider growth in the perceived importance of lexis as a language system, promoted by research such as Hill (2000), who found that up to 70% of everything we read or hear is part of some kind of fixed expression (see also Meara, 1980 and Wray, 2002, 2013, 2018). James (1998) offers five good reasons why LEA should be undertaken:

- 1) Lexis is taking a more central role in language study (morphological aspects of words, multi-word lexical units including, idioms and their centrality to expert-like fluency).
- 2) Language learners themselves place great importance on learning the lexical system of a language, perhaps more so than grammar and pronunciation.
- 3) Lexical errors have been proved to account for the greatest percentage of errors in student output.
- 4) Those

with English as a first language consider lexical errors to be the most disruptive and irritating of all error types. 5) Lexis carries a much greater communicative function than other language systems.

Further inspiration for completing this work came from the researcher's interest in lexis as the most important language system (Lewis, 1993) and a statement from Hemchua and Schmitt (2006); the authors predicted, after analysing the lexical errors they found in the compositions of 20 Thai learners, that 'similar findings would be found for different L1 groups'. If this were true, it would suggest that all learners, whatever their L1, would make similar lexical errors in their compositions. It would therefore be worthwhile ascertaining whether this is true or not; there are probably patterns of similarity, but it seems likely also systematic differences. Therefore, it would be very useful to establish precisely what type of errors different types of learners are making and their frequency. This information would be of great interest to SLA researchers, teachers and learners. There have been various attempts to produce frameworks for LEA (e.g., Dušková, 1969; Corder, 1973; Zimmerman, 1986; Engber, 1995; James, 1998 and Hemchua and Schmitt, 2006), many of which represent milestones in understanding in this area. However, none of these authors completed replication studies to ascertain agreement between raters when using their frameworks, so it is not known how similar findings will be when used by different analysts on the same composition data. Furthermore, an analysis of the various frameworks above points to strengths and weaknesses in each. Therefore, in order to serve the various groups mentioned above, a single framework that incorporates the advantages and eliminates the weaknesses of these frameworks should be produced.

1.2.1 Uses, users and contexts for the LEA tool

If such a framework could be devised to identify, count and categorise the lexical errors made in the written compositions of learners, this very detailed feedback on what a learner does not do well with lexis could be used to refine lexical accuracy by several different groups, such as learners, teachers, markers, materials writers and second language researchers:

1.2.1.1 Learners If the metalanguage used was not too far above the learners' comprehension, the analysis results would be of great interest to the more motivated language learners themselves. They would be able to see the areas in

which they were making errors and take steps to eradicate these, thereby also enhancing learner autonomy.

1.2.1.2 Teachers In contexts where there is a greater demand for lexical accuracy, such as in any high-stakes exam preparation classes, the analysis will help learners and teachers to understand how to improve their lexical resource. The results will be of interest to their language tutors, as areas to be improved will be uncovered. Tutors will also be able to see where language taught is being used well and where it is not, thereby possibly confirming useful teaching methods and identifying less effective ones. It could also signpost language for remedial teaching. This would be useful in courses taught both in countries where English is the dominant language and in others. It is perhaps more useful in EAP contexts where students are being prepared to write with great specificity in their writing, rather than in contexts where fluency is the main goal. In ESOL classes in the UK, for example, the tool's use may be of less importance. In terms of use in countries that have their own variety of English with their own levels of acceptance of deviation from British English, for example, the tool could be used by experts in those varieties and modified accordingly so that acceptable uses in that context are not penalised for not conforming to an unused variety. The tool could be used on an individual student's writing to provide detailed feedback by identifying each error and a correction and by categorising them so that the student could see the most common types and amounts of errors made. The teacher could supply guidance to avoid these kinds of errors in the future. As individual feedback may be too time consuming for the busy practising teacher, LEA could be performed on the essays of the whole class, and types and totals noted for whole class remedial work.

LEA could also be used as part of a diagnosis of lexical ability before teaching commences, thereby focussing what to teach.

1.2.1.3 Markers The tool could be used in assessment to standardise markers' scores for 'lexical resource'. At the moment, I feel there is insufficient guidance for markers working on IELTS scripts. The band descriptors for this criterion are open to interpretation. Trained lexical error analysts may be more likely to gain inter-rater reliability.

1.2.1.4 Materials writers will also be interested in the results of LEA, as they may influence specific areas to be taught in coursebooks. Publishers could also use it to produce and market lexically-graded books

1.2.1.5 Second Language Researchers will find LEA results of interest as longitudinal case studies can chart the development of lexical acquisition over time. SLA researchers could also compare lexical accuracy with other variables, such as an overall writing score, as Llach (2012) did, for example.

However, there are natural limitations to how the tool could be used. It should not be seen as the sole feedback mechanism or for a test of student writing as it cannot be used to comment on punctuation, whole text organisation or how well a piece of writing answers an essay brief, for example, nor can it be used to comment on what the student has done well (but see Section 5.2 for a discussion on how the tool can be adapted to analyse other areas of language production). Given that spoken output contains a higher percentage of 'slips of the tongue' than written output, its usefulness for analysing spoken accuracy is no greater than other methods.

1.3 Organisation of the dissertation

Chapter one provides research aims, background and justification for the study and signposting for the dissertation as a whole.

Chapter two is a literature review that initially explores EA in general and describes its rise to popularity and fall from grace in the 1960s and 1970s. The chapter provides a close examination of the reasons for its decline in popularity, and evaluates counter-arguments. The chapter examines the importance of lexis in language and language education, and provides a justification for LEA, based on the rise of the centrality of lexis in language teaching and learning and the blurred boundaries between lexis and grammar. It also explores various LEA studies and critiques the different frameworks for their perceived potential ease of use and depth of analysis.

Chapter three, Study 1, (published as Picot, 2017) is a replication LEA study, using Hemchua and Schmitt's (2006) framework and guidance. They sought to establish the type and frequency of lexical errors made by 20 Thai university students in their written compositions. A replication study was necessary to investigate one of the more recent and comprehensive frameworks for LEA, and whether LEA was a worthwhile and useful area for development. Hemchua and

20

Schmitt's framework was based on work by Leech (1981), Laufer (1991) and James (1998). As mentioned above, the authors stated that other L1 groups would make similar LEs to the ones they found in their study. If true, this would be of great interest to SLA researchers, teachers and learners. Therefore, 20 compositions of a similar length were collected from Greek learners of a similar age and proficiency level, and the Hemchua and Schmitt (2006) framework was used to analyse their work. It was found that there were indeed many similarities in the type and frequency of lexical errors made, confirming, to some extent, the authors' hypothesis above. However, perhaps the dissimilarities between the two studies' findings could be explained by the difficulties in error identification and categorisation experienced when using the framework. More guidance for the former was required and a better system for categorisation was required: one which would not provide dual categorisation issues.

Chapter four, Study 2 - Given the similarity of results in the previous study, it was clear that LEA has much potential as a valuable tool for SLA researchers, teachers and learners. However, it should be as straightforward to conduct as possible and should provide a detailed range of results for it to be of as much practical value as possible. Therefore, given the issues discussed in Chapter Two in terms of problems of error identification and categorisation, a new improved framework was required. To this end, the main LEA frameworks published in the last 60 years were examined for their ease of use and depth of analysis. Six quite different frameworks (Dušková, 1969; Corder, 1973; Zimmerman, 1986; Engber, 1995; James, 1998 and Hemchua and Schmitt, 2006) were selected for use in the study. Selection was based on the studies' perceived usefulness and different approaches to LEA. To enable comparison of findings across the analyses, five of the essays used in the previous chapter's study were used. The advantages and disadvantages of each system were noted so that a new, improved LEA framework could be produced (NewLEAF Version 1).

Chapter five describes the new framework (NewLEAF1) and offers guidance for error identification and categorisation, two aspects that were found to be problematic in the literature review and in the previous two studies. The guidance contains clearer rules for error counting and the separation of grammatical and lexical errors. The taxonomy in the new framework is organised into formal and semantic errors and employs several of Hemchua and Schmitt's (2006) error types, as well as those from Zimmermann (1986). It incorporates James' (1998)

suggestion for a dual-axis framework and avoids speculative issues of cause of error by removing categories, such as first language (L1) interference. It avoids further unscientific issues, such as deciding on the seriousness of error as incoherence errors are surely the more serious as they cause breakdowns in communication, and further ranking of the seriousness of error types would be subjective and context-dependent, and therefore, best left to the individual teacher, learner or analyst to decide upon. Finally, the framework is tested (Study 3) using the same essay data that was used in Chapters 3 and 4 (Studies 1 and 2), for comparative purposes. The results are compared with those obtained using other frameworks.

Chapter six, Study 4, sought to examine the degree of similarity of analysis that was found when the guidance and framework in NewLEAF1 were used by seven highly experienced and highly qualified English expert university EAP teachers. The participants were given one digitised script produced by an English language learner. They were put into pairs and audio recorded whilst discussing their thinking and decision-making whilst completing the following two tasks. They were first asked to identify all errors by highlighting the digitised scripts and then decide whether they were lexical or grammatical in nature, or other. The second task was to categorise 37 extracted lexical errors using the framework and allocate a confidence score of 1-3 to show how sure they were about their categorisations. The annotated scripts and completed charts were then collected and compared to ascertain the similarity of results and any discrepancies between them. The recordings were analysed to understand the decision-making behind their results. The analysts were subsequently interviewed to ask them about any issues they experienced whilst performing the LEA, what they thought in general about the process and concept, how they might use LEA in their work and whether they had any recommendations for changes to the framework and guidance. These data were used to produce NewLEAF2, which would also require testing.

Chapter seven, Study 5, trials NewLEAF2 with 41 participants. It repeats some aspects of the first study, with a focus on quantitative elements. The same essay and procedures were used. The framework and guidance should be useful to teachers and linguists with various backgrounds, so these participants were a mixture of TESOL and Linguistics university staff and students. One third had English as a second language. Again, the results were analysed for agreement.

Identification and dual categorisation issues were analysed and a simplified version was produced, NewLEAF3.

Chapter eight, Study 6, is the application of NewLEAF3. The original Greek data from Study 1 were reused. This enabled some comparison of findings between the current study and Study 1. It was found that there were no uncategorisable errors, and in terms of possible dual categorisation issues, there was only a small percentage of confidence scores of 2.

Chapter nine provides a discussion, conclusion and summary of the dissertation, its limitations and possible future directions for LEA research.

Chapter Two: Literature Review

2.0 Introduction

This literature review will explore topics that are central to the following line of argument: LEA is an under-researched aspect of applied linguistics and therefore warrants investigation into its potential for shedding light onto the types and frequency of LEs made by English language learners. This information will be of interest to materials writers, teachers, SLA researchers and not least, learners themselves. An effective and usable framework needs to be based on a clear understanding of lexis and LE: it needs to take into account the strengths and weaknesses of existing frameworks, to strike a satisfactory balance between ease of use and depth of analysis, and to produce similar results when used by different people.

The literature review starts with a definition of 'error' in language learning, a brief discussion of the views on errors and whether they should be corrected in language teaching.

Following this is a discussion of the value of EA and its applications. The position put forward is that despite the conflicting views on the necessity and efficacy of EA, it makes an important contribution to helping teachers and learners understand the issues that they are facing and that it can inform areas for remedial teaching.

Next, the history of EA is charted. The prevailing theory of SLA before the rise of EA, i.e. Contrastive Analysis, is explained and reasons are presented for its demise. The reasons for the rise in the popularity of EA are examined.

The process of EA is more closely examined next, alongside the issues that led to its fall from popularity. The traditional stages in EA are collection, identification, description, explanation and evaluation (Corder, 1967). Next, the different ways that errors have been categorised are explained. The limitations are then summarised, a rebuttal is offered and suggestions made to minimise these limitations. A line of argument is put forward to support the value of EA.

Then the focus turns to LEA. There is an explanation of the importance of lexis in communication and its centrality in language learning and teaching. There is a discussion of what constitutes lexis and grammar: how it is problematic to separate the two and a discussion of the concept of lexicogrammar. Following this is a case for LEA and a discussion of definitions of LEs in writing and an examination of the various types of LEs in the rather limited body of research into the topic. This is done by analysing 12 of the various frameworks that have been

used to categorise LEs in the last 60 years or so. These frameworks are critiqued with reference to potential ease of use and depth of results.

Finally, there is a summary and a conclusion to the literature review.

2.1. Definition of errors in language learning and teaching

Lott (1985:258) states that 'to be able to categorise errors without great difficulty, it is essential to have precise definitions'. This section examines various definitions of error and proposes one that will be used throughout this dissertation.

It is interesting to note the various definitions of error that have emerged in the literature since Pit Corder first started writing about them in 1967. He stated that originally, errors were 'dismissed as a matter of no particular importance, as possibly annoying, distracting, but inevitable by-products of the process of learning a language'. (1967:147). Dulay *et al.*, (1982:183) defined them as 'the flawed side of learner speech or writing [...] those components of conversation or composition which deviate from some selected norm of authentic language performance'. Ellis (1994) also defines errors as deviations from the norms of the target language. George (1972:2) circumnavigates this problematic area of reference to the norm and error identification by simply stating that an error is an unwanted form, and adds that if a speaker of a different variety of English (e.g., Indian English) makes an utterance in that variety that is deviant from standard English, it should not be considered an error.

So it seems there is agreement that errors are deviations from norms: different from what an expert user of the language would expect. However, those with English as a first language also commit errors in output, especially when speaking. For example, the language output of a very young child is peppered with non-standard usages in vocabulary, such as overextension (Clark and Clark, 1977). This led Lennon (1991:182) to define an error as 'a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's native speaker [*sic*] counterpart'. However, there is still the issue here of whether this relates to standard English or British English, for example. James (1998:64) endorses this definition as it 'sidesteps the problem of semantic and formal intention'. However,

this appears to immediately introduce doubt and confusion into the process from the start. An error analyst might spot the error '*There *is many cars on the road today.*' and be tempted to disregard it as a common error made by those with English as a first language. However, this dismissal will not help us in the systematic identification of errors in error analyses. Ignoring it will also not help the learner that uttered it, and it is still an error from a prescriptivist point of view. Therefore, for our purposes, a combination of Lennon's (1991) and George's (1972) definitions of error will be used: in this thesis, an error is defined as a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's expert language user counterpart when communicating in the same variety of English.

2.1.1 Justification for error correction

Not only is the definition of an error problematic, but there is also the issue of whether and how teachers should deal with them. Linguistic opinions on correctness can be broadly categorised as prescriptive and descriptive. Prescriptivists typically focus on correctness, on accurately following a notional set of rules. This view is followed by some language teachers, students and exam boards, as evidenced by the latter's published marking criteria. A descriptivist account, in contrast, seeks to describe language use without passing judgement on correctness. Use of the term 'error', unfortunately, views it as a negative feature of a learner's interlanguage ('interlanguage' here follows Selinker's (1972:35) definition as 'a separate linguistic system based on the observable output which results from a learner's attempted production of a target language norm'.) Pit Corder (1967) coined the term 'idiosyncratic dialects' when referring to learner output or interlanguage. Recently, following the growing rise in popularity of task-based learning (Ellis, 2005 and Long, 2014), the focus in language teaching literature has changed from accuracy to the ability to get one's message across. However, it is this author's opinion that there is still and should still be a requirement for accuracy in student output. Indeed, LEA is wholly compatible with a task-based learning approach to language instruction, as LEA can, and perhaps should, be used in the final stage of Willis' (1996) task cycle, where learner output is analysed for accuracy. Further, given the high-stakes nature of the IELTS and other similar tests, where lexis is assessed for accuracy and range, a focus on

LEA can be wholly justified. Indeed, students themselves still view language learning in terms of 'correct' and 'incorrect' and many want errors to be corrected (Papangkorn, 2015).

However, there is still the issue of whether they should be corrected. Some authors, most notably Krashen and Terrell (1983) with their 'Natural Approach' take a *laissez-faire* view of error correction. They see language learning as learning to transmit meaning and if the transmission is successful, correcting errors is of little importance. They believed that accuracy is subordinate to fluency, and learners would become more accurate in time, regardless of how they were instructed, as long as they were exposed to rich input. There should only be a focus on errors if they interfere with the transmission of this communication. Whilst this final point will appeal to teachers who follow a communicative approach to language learning, Krashen and Terrell's (1983) assumption that accuracy will follow fluency has been questioned in the literature (e.g., by Housen *et al*, 2012) and the validity of their claims relate more to natural acquisition, where languages are picked up through exposure and incidental learning, rather than classroom learning which is explicit and often form focussed. Also, some teachers may believe that the correction of learner errors may lead to a reduction of self-confidence for those learners and heightened anxiety levels, which may 'raise the learners' affective filter' (Krashen, 1985), making language learning more stressful and therefore more difficult.

However, there is more substantive evidence for the opposing viewpoint: that error correction is a valuable and important feature of language instruction. Although learners can indeed progress well with comprehensible input, a lack of error correction may lead to fluent but highly ungrammatical or inaccurate communicators (Richards, 2002). More modern thinking, since Krashen and Terrell's (1983) claims, states that taking a more form-focussed approach is more effective (Willis, 1996). Task-based language teaching, which encompasses a focus on form in a post-task analysis of learner errors, has been shown to be effective in eradicating those errors (Ellis, 2003, 2016; Long and Robinson, 1998 and Long, 2014). Feedback is an important step in mapping form against meaning for the learners (Lyster and Ranta, 1997; Li, 2010 and 2013 and Mackey, 2013). 'Most practitioners would agree that correcting all errors is not useful, realistic or possible, but that appropriate corrective feedback is likely to facilitate L2

developmental processes' (Mackey, *et al*, 2016). This is an interesting statement that warrants further investigation. Firstly, one may need to identify all of the errors before deciding which are the more serious and deserve attention, and secondly because it may be possible to identify all of the errors in a student's composition. This viewpoint supports EA.

My experience as a language teacher also suggests that when done sensitively, error correction is not only expected by motivated learners, but also a valued and gratefully received feature of instruction, and allows learners to notice the difference between their own interlanguage and target norms. This is widely agreed in the literature as an important step in SLA (e.g., Schmidt, 2001). Not only is it important for learners to understand the types and quantities of errors that they are making, but an understanding in these areas will help teachers to identify where instruction has been successful and where it has not, which will inform remedial teaching. This knowledge should also inform materials writers. Finally, an understanding of the types of errors that learners make will also inform SLA research. These are the lines of reasoning that led to the popularity and rise of EA in the 1960s and 1970s.

2.2 Error analysis introduction

2.2.1 Definition and purposes

Broadly speaking, 'EA consists of a set of procedures for identifying, describing, and explaining language learner errors' (Ellis and Barkhuizen, 2005). James (1998:1) defines EA as 'the process of determining the incidence, nature, causes, and consequences of unsuccessful language' and Erdogan (2005:263) describes it as a method of assessing the learners' performance 'in terms of the cognitive processes they make use of in recognising or coding the input they receive from the target language'. According to Dulay *et al.* (1982:138), it 'indicates to teachers and curriculum developers which part of the target language students have most difficulty producing correctly, and which error types detract most from a learner's ability to communicate'. More specifically, and according to its creator, Pit Corder (1967), EA investigates the types and causes of language errors through a five-step process, often via learners' written compositions. This involves collecting samples of learner language, identifying the errors (includes differentiation between errors and mistakes), describing the errors (systematic categorisation),

explaining the errors and evaluating (for seriousness) and/or correcting the errors (see Section 2.4 for more information on the processes involved).

Furthermore, according to Dulay *et al.* (1982:138) EA is done 'to provide data from which inferences about the nature of the language learning process can be made'. However, James (1998:1) quotes Cook (1993:22) as saying that EA is 'a methodology for dealing with data rather than a theory of acquisition'. This last point may explain some of the dissatisfaction with EA. It was indeed once proposed as the predominant theory of acquisition (James, 1998), as one would be able to predict the L1 transfer errors that any student of any given L1 may make and therefore how they would acquire the L2. However, it is flawed as a singular theory of acquisition because it is not comprehensive and cannot explain or take into account other types of errors, such as universal developmental errors that are not caused by L1 influence or teacher induced/materials induced errors (Richards, 1971).

That said, by utilising a Contrastive Analysis (CA) (see Section 2.3.1) approach, Swan and Smith published a seminal and very valuable set of articles in their 2001 book, *Learner English: A Teacher's Guide to Interference and other Problems*, which informs teachers about common errors made by their L1 groups. The book contains chapters on the major world languages and predicts probable problems, based on L1 interference, which teachers of students of those L1 groups should be aware of. The popularity of this book today, indicated by a high number of five-star reviews on Amazon, and now in its second edition, indicates its usefulness. If the teacher could make her students aware of these errors, it would massively increase the accuracy of their output.

It is my opinion that EA as a theory of language acquisition is of little value, but perceived issues with its various processes and procedures (see Section 2.4.7) can be mitigated or overcome, and it should still be seen as a useful tool in the teacher's toolkit.

2.2.2 The need for EA

There are several reasons why this potentially problematic and time-consuming process should be carried out, but broadly, they can be divided into academic/research-based reasons, which inform studies into SLA, and pedagogic

reasons, which can be used to assist learner proficiency through teaching and materials writing. For both we need systematic, objective and reliable methods and procedures to measure the writing ability of L2 learners.

Interestingly, James (1998:8-9), following Schmidt's (1990) noticing hypothesis, where it is stated that learners who notice differences between their own interlanguage and a target norm learn best, states that learners often conduct their own internal EA when they compare their interlanguage with the target language when they notice the difference between their own output and that of the target language, and restructure their hypotheses accordingly. EA may therefore be considered a procedure that builds on and enhances the learners' own actions.

2.2.2.1 Academic/ SLA research

Some teachers may see errors as a result of failures in teaching and learning (Richards and Rogers, 2001), but this could be considered an out-dated viewpoint based on a fundamentally behaviourist approach. Errors are often the result of learners trying to convey meaning with language that they have not yet mastered (Long, 1996; Ellis, Basturkmen and Lowen, 2001 and Swain, 2005) and should be viewed positively. Given the complexity of language learning, they are inevitable and can tell us much about the process of SLA and how to remedy the errors and improve performance. Learners might have an 'inbuilt syllabus' which determines the order in which grammar is learned or acquired' (Buyl and Housen, 2015).

Studying errors might provide clues to the order of competence in these areas (Corder, 1967). Furthermore, errors provide 'evidence of how languages are learned and serve as devices by which the learner discovers the rules of the target language' (Corder, 1967). EA 'helps us to understand how SLA proceeds and where it does not', its stages and what processes are involved, and where in terms of acquisition/learning the learner is (Llach, 2011: xi-xiii). Many SLA studies (e.g., Llach, 2011) attempt to measure accuracy in language output and seek correlation with other variables. Having a more reliable and valid framework for EA will enhance the accuracy of such studies.

2.2.2.2 Pedagogic uses

EA can be used to investigate English Language teaching and learning problems, not only for second language learners, but also for first language learners (James, 1998). It can suggest what has been learned and what has not (Corder, 1967). For this, we also need error evaluation (Ellis and Barkhuizen, 2005) to prioritise what

should be taught remedially. Errors can point out methodological shortcomings in teaching as well as learning problem areas. Therefore, EA should ideally precede teaching students that are new to a course (Llach, 2011: xi-xiii). However, for beginners at least, some teaching needs to have taken place before learners can supply a sample for analysis: in such cases, perhaps teach/analyse/teach is a better model. Errors 'serve a learning purpose by acting as devices by which learners can discover the rules of the target language, i.e., by obtaining feedback on their errors' (Ellis and Barkhuizen, 2005). EA can be used to identify successful teaching and learning strategies and can be used to design remedial teaching materials.

Etherton (1977) provides the most comprehensive rationale for conducting EA, identifying 11 reasons. Of those not already mentioned above, he argues that no textbook, unless it is designed for a specific group of learners, will be able to include all the materials the class needs, and therefore, EA should be performed on the writing of a cohort. It can highlight common weaknesses and identify both language which is too difficult and for which teaching should be deferred, and language items which have been avoided and should therefore be (re-) taught. However, students may avoid a structure because they do not like it or feel comfortable using it, or perhaps that structure does not exist in their L1. It does not necessarily mean that they have not learned this structure: another inherent weakness of EA. However, it could be argued that EA will identify aspects of language that are required for communication and should therefore be taught. It is also a form of self-education where a teacher can familiarise herself with common issues amongst her students or where teaching has been unsuccessful.

2.2.3 Other uses of EA

According to Doolan and Miller (2012), EA has been widely used for making judgements on gauging proficiency levels (Bardovi-Harlig and Bofman, 1989 and Ho-Peng, 1983) and measuring the gains made through writing instruction (e.g., Ishikawa, 1995 and Shaw and Liu, 1998). Larsen-Freeman and Strom (1977) stated the importance of including errors when evaluating writing, as they demonstrate gaps in syntactic control. Because EA has been an effective approach for distinguishing groups of L2 student writing, it may also differentiate Generation 1.5, L2 and L1 texts (Doolan and Miller, 2012). Felix (1981) and Pavesi (1986) used EA to compare language used by instructed and naturalised learners.

Bardovi-Harlig and Bofman (1989) also investigated differences between learners who passed a university placement exam and some who did not. 'Major differences were found in the number of morphological and LEs, rather than syntactical' ones (they found different kinds of errors for different levels). Santos (1988) carried out error evaluation to investigate linguistic claims regarding markedness. Ferris (2002) supported Hendrickson's (1978) appeal for researchers to investigate error as a basis of deciding what to teach next, especially in remedial approaches to teaching writing (Ellis 2008). It has also been used to test for accuracy in studies of e.g., task design (Foster and Skehan, 1996).

James (1998:56-61) outlines several other areas that could make use of EA: monitoring language change (what was once seen as an error may now be seen as acceptable usage); monitoring language loss; highlighting deviance in literature (artistic license permits original phrases); and using EA in instructional science (where rules concerning formulae and language are inferred incorrectly). Furthermore, according to James (1998:xi), apart from SLA studies, EA has been used by those interested in 'mother-tongue literacy, oracy and writing assessment, language disorders and therapy work and the growing field of forensic linguistics'.

2.2.4 Computer-aided EA

EA has been reborn in computer-based analyses of learner language, but according to Ellis (2008), not all computer-based analyses of learner corpora have utilised EA effectively. We are yet to see a system whereby we could satisfactorily analyse error with computers. Unfortunately, software is not yet sophisticated enough to identify all possible errors. Neither is there a suitably sophisticated framework for dealing comprehensively with EA for human analysis: hence the focus of this thesis.

Corpora have been collected for the International Corpus of Learner English (ICLE) project, where Granger (1998) originally used a tagging system for each instance of error type (Grammar, lexical, lexico-grammatical, formal, register, syntax and style) and then for sub-types. E.g., Grammatical verb error or grammatical auxiliary verb error. This is an area that appears to be growing in popularity. Various organisations have collected large learner corpora (e.g., Open Cambridge Learner Corpus) and can use software to analyse them for error. However, for most of these systems, before any computational analysis can be

carried out, the data has to be encoded, marked up, tagged and parsed by a human analyst (which probably takes a similar amount of time to conducting a manual EA). In other words, manual EA has to be carried out before the digitised data can be used. Therefore, because EA is a preliminary stage to computerised tagging, the focus of this thesis will remain on manual EA and the pursuit of a useful framework, which could later be applied to computerised investigations. The Standard Speaking Test Corpus (SSTC) project collected oral samples and placed students into one of nine proficiency levels. It used a tag editor to facilitate the tagging of errors (Tono *et al*, 2001). Under error tagging, three broad types of error were identified (omissions, replacements and insertions) and 42 error categories were distinguished. The error tag provided part of speech, the rule that had been transgressed and the corrected form. e.g., **I belong to two baseball <n_num crr = "teams"> team' <n_num>*. Once a corpus has been electronically tagged, a detailed analysis can be carried out, identifying frequency of different error types, allowing analysis of large corpora and comparison of errors produced by different learners' proficiency, age, gender, etc. Unfortunately, this process will be very time-consuming for the practicing EFL teacher who wishes to establish the weak areas of her students' writing. A new framework for the analysis of LEs could be used with this electronic tagging. However, there have been some useful advances in this area: Answers to problems of use of forms have been proposed by Granger and Tyson (1996) who used computer-based language corpora from L1 users to make comparisons of suspect utterances with baseline data, making awkward expression use more easily investigated.

Götz (2015) used computer aided grammatical EA to propose a system whereby quantitative analysis can be used to complement the qualitative descriptor bands used by the CEFR to allocate learners to proficiency bands. Despite Granger's (2002) and Lüdeling *et al*'s (2005) compelling claims that error-annotated corpora are a valuable tool which can be used to understand SLA, Kreyer (2015) states that one of the issues with computer-aided EA is that there is still a lack of comprehensively annotated learner corpora. Granger (2008:266) states that issues relating to the steps of computer-aided EA (identification, correction and annotation) are still beset by a series of problems and it is doubtful that a perfect system for annotation will ever be found. This calls for some of the issues with EA to be resolved before further advances can be made. These issues resemble those found in more traditional EA processes. Kreyer (2015) states that the

greatest issue with computational error tagging is issues in deciding what the reformulation of the error should be: an error cannot be categorised if there are different possibilities for the target form. Having seen various categories (or tagsets) for the categorisation of errors in learner corpora, it seems to me that the greatest issue is the lack of categories for analysis that they offer and subsequent lack of depth of analysis. Despite these issues, some progress in computer-aided EA has been made by Gamallo *et al* (2015) in terms of automatic error tagging. They developed a system entitled Avalingua, a program designed for Portuguese L2 learners of Galician. It uses Natural Language Processing and knowledge-rich linguistic resources, such as lists of common error types and linguistic rules for automatic error detection (spelling, syntactical and LEs). However, whilst the authors state that it could be developed for English, the number of categories for LE categorisation seems quite low, and whilst it can identify and classify LEs based on rule-bound concepts, such as false friends (between Portuguese and Galician) and affixation, it does not take into account errors of collocation or connotation, amongst other possibilities. It is expected that this field and systems such as Avalingua will advance in the coming years, but its success will depend on a reliable system of error identification and classification, which must first be trialled with human raters.

2.3 A history of EA

2.3.1 Before EA: Contrastive analysis and its demise

The prevailing theory of SLA in the 1950s and early 1960s was broadly behaviourist (Richards and Rogers, 2001). Proponents viewed errors as problems that had to be dealt with: they were evidence that learning and teaching had not taken place smoothly. If applied linguists could investigate the differences between the first and second languages, certain errors could be predicted (up to 30% of all errors, according to James, 1998:4), explained and therefore avoided with the design of instructional materials. The thinking was that if there was similarity between associated structures, these structures will be easily acquired, but if there were great differences, these would require more attention. This was Contrastive Analysis (CA: also known as Contrastive Analysis Hypothesis or Contrastive Linguistic Analysis), the prevailing method of studying SLA in the 1960s (James, 1998:4). This approach was advocated by Schachter and Celce-Murcia (1977:441), who stated, ‘...a systematic comparison of the source language and the target language at all levels of structure will generate predictions about the

areas of learning difficulty in the target language for speakers of the source language....consequently, the best teaching materials will emphasize those features of the target language that differ markedly from corresponding features of the source language...'. However, as a theory of language acquisition, CA fell from favour, as it was associated with an outdated, structuralist model of language description and a behaviourist theory of learning. Furthermore, Schachter (1992) found that some of the structural areas where the first and second language were similar were not easily mastered and not all of the main areas where structures differed were difficult for learners. CA was also proved insufficient as a theory of SLA: early error analyses exposed that not all the expected errors were made by learners and many errors were attributable not to interference from the learners' L1, but to the complexity of the L2, faulty inferences made about the rules of L2 or lack of exposure to the L2 (Harshbarger & Gordon, 2003:44). It was also felt to be too subjective in terms of what the analyst felt was an error, and lacked predictive power (Al Kresheh, 2015). These were the main reasons for the demise of CA. Despite the fact that it attracted much criticism in the late 1960s, CA was the early forerunner for the Swan and Smith's illuminating 2001 book, *Learner English: a teacher's guide to interference and other problems*, mentioned in Section 2.2.1 above.

2.3.2 The rise of EA

As mentioned above, EA in SLA was first discussed in publications in 1967 by Stephen Pit Corder and colleagues. It was viewed as an alternative, a reaction to the shortcomings of CA. In his milestone article, *The Significance of Learner Errors* (1967), he stated that second language errors are of interest as they reflect the underlying linguistic rules of the learner's interlanguage. Understanding how they understand the language would help us teach them. Dulay and Burt, in 1974, stated that EA was an improvement and CA was too restrictive. The main difference between the two approaches were a focus on potential errors (CA) and actual ones (EA). Another was in their underlying assumptions about SLA. EA showed that the internal processes of learning a language were psycholinguistic in nature (learners' internal mental processes were responsible for many of the errors made, not L1 or shortcomings in behaviourism). Analysis of these errors help applied linguists understand how a learner develops hypotheses about the structure of the L2. Schachter & Celce-Murcia (1971:142) stated that EA

demonstrates that not all systematic errors are due to L1 interference and that many errors show that there is a much more complex learning process at hand: one where 'the learner is an active participant' in the formulation and revision of language hypotheses. Several researchers have contributed to the understanding of SLA through EA (Corder, 1967, 1971, 1974, 1981; Dušková, 1969; Nemser, 1971; Richards, 1971, 1976; Selinker, 1972; Politzer & Ramirez, 1973; Stenson & Schuman, 1974; Ghadessy, 1980; James, 1998). EA became very popular between 1967 and 1979. This gave rise to a flurry of research papers investigating the errors made by different L1 groups, such as Spanish, German and French. These were gathered and a comprehensive bibliography was published by Spillner in 1991, and an updated edition was published in 2017.

2.4 The process of EA

This section describes, in detail, the stages of error correction and their associated issues. According to Corder (1974 in Ellis, 2008), the following are the five steps that must be followed to effectively analyse errors:

2.4.1 Collection of errors

A variety of sample types can be used: different mediums (written or oral), different genres (e.g., essay or letter). These may produce different types of errors. LoCoco, (1976) established that there were differences between the amount and type of errors from free compositions, translations and picture composition. Samples may be drawn from different levels of learners with different language learning histories (language learned or acquired) and different exposure to the target language. The type and number of errors will also vary due to whether the samples are natural, spontaneous, careful or elicited, and how familiar the learners are with the topic (Ellis 2008:47). Previous studies did not attend to these differences, making interpretation and replication problematic. Ellis (1994:49) states that one could also consider the following factors when seeking a sample: language (medium: oral or written, genre (letter or essay), content (the topic that the learner is communicating about) and the learners themselves (proficiency level, their L1 and language learning experience: classroom or naturalistic). All these factors will have some impact on the errors that are or are not produced. Ellis (2008) also states that natural samples of data are better, but hard to obtain. Corder (1973) argued for elicited data and suggested two types: clinical and experimental. 'Clinical elicitation involves getting the informant to produce data of

any sort, for example by means of general interview or writing a composition. Experimental elicitation involves the use of special instruments to elicit data containing the linguistic features such as a series of pictures, which had been designed to elicit specific features' (for example past simple narrative). James (1998:19-24) classifies the types of possible data collection as observational (classroom activities, such as role-play, simulations, and information gap activities) and experimental (where, for the specific purpose of EA, target forms are elicited with special instruments. He adds a third category: that of introspection, where a learner is asked to explain their language production choices. This allows for a sample as well as an account of why they said what they said. Learner diaries are also a valuable source. However, researchers do not always have the choice of sample type, as they are not always able to intervene, and may have to work with what the learners are producing anyway (for homework or exams for example). Teachers seem best placed to collect learner samples.

It seems logical that if an analyst wishes to observe the state of a learner's interlanguage, then written samples are more desirable, as with writing, a slower form of output, the learner has more opportunity in the composition process to monitor for, and self-correct, mistakes. Spoken output is faster as it is produced under the processing conditions of interlocutor and real-time constraints (Bygate, 1987), and therefore more prone to mistake production. Of all the stages in EA, collection seems to be the least problematic. The next, identification, has many related issues, not least differentiating between errors and mistakes.

2.4.2 Identification of errors

Errors Vs Mistakes

It may be a better use of the learner's and analyst's time if they were to focus on true errors and not on slips (mistakes). If one is not to waste time or be misled when considering what the learner does not know, the difference between these two should be established. This, however, can be problematic. Also, students may wish to have their slips pointed out.

Mistakes are 'random deviations unrelated to any system and represent the same type of performance or writing of a native speaker' (Falk, 1978: 360). Corder (1973: 258) asserts that they are 'the outcome of non-linguistic factors like memory limitation, strong emotion, fatigue, lack of concentration, etc'. Chomsky (1965) helps us to understand the difference by making a distinction between competence error (mistake) and performance error (error). One may repeatedly see the same deviation from the norm in a piece of writing (a competence error). Or one may see a structure or word used correctly most of the time, but then the same language used incorrectly on one occasion (a performance error or mistake, demonstrating a U-shaped development curve). Here, it is possible that a learner is aware of the correct form but produces an accidental slip when producing language. The latter could be due to inattention, memory-lapse or fatigue. Mistakes can be self-corrected by the learner and errors cannot. Errors are systematic and show the learner's understanding of the L2. Corder (1981) also makes the same distinction, naming errors systematic and mistakes non-systematic. Systematic errors show the learner's understanding of the language. These should be the focus of EA, and performance or non-systematic errors should be dismissed (Corder, 1981:10). However, being sure which is which can be problematic. Corder (*ibid*) argues that EA assumes that competence is homogeneous rather than variable: i.e. sometimes learners misuse a target form, and at other times, they do not. EA has traditionally ignored this problem of variability.

Ellis (1985) states that there are two methods of differentiating between errors and mistakes. Firstly, as mentioned above, if a learner consistently produces an erroneous form, it is an error. If he/she is sometimes correct with this form, it is a mistake. Therefore, Corder states that successful EAs can be carried out only when the intended meaning is clear. This is a further issue in error identification: the analyst cannot always be sure of the intended meaning of the error. Therefore, according to Corder (1967), a second method, a plausible interpretation of the learner's meaning must be made. Corder (1971 and 1974), acknowledging the importance of interpretation, adds a third when he proposed an elaborate error identification procedure:

- a) First, where the writing is apparently error-free, it must be considered a 'normal interpretation or not apparently erroneous', but it could be right by chance.
- b) If an error is found, one must seek an 'authoritative interpretation', which involves asking the learner what was meant and making an 'authoritative reconstruction'. However, learners' retrospective accounts of intended meaning are often not reliable (Van Els *et al*, 1984 in Ellis, 2008)
- c) If the learner is unavailable, a 'plausible interpretation' could be gained by examining the context of the error or by translating it into the learner's L1. However, the analyst may not always be confidently bilingual.

Consulting the learner might seem to be the best way of ascertaining meaning, but would be very time consuming or impossible at all times. Therefore, plausible interpretation would be the most efficient way.

Lennon (1991) discusses problems in identifying errors, especially spoken ones (local and global), and concludes that 'difficulties in identifying errors will increase with the informality of the discourse' (1991:182). This argues the case for written samples of language, possibly English for Academic Purposes (EAP) scripts which should be formal in style and therefore any use of informal language is clearly an error. Therefore, using the word 'kids' in a formal essay would indeed be considered an error. These same problems led Corder to decide that 'every sentence is to be regarded as idiosyncratic [possibly erroneous] until proven otherwise.' (1974:21).

In Lennon's (1991a) study of the spoken errors of advanced German learners of English, approximately 30% were considered borderline cases. Consensus of opinion could still not be reached after he submitted these for consideration to a panel of nine educated expert English speakers. This finding represents a serious barrier to successful EA and prompted him to establish the concepts of domain and extent to help with identification. In a seminal paper in 1991b, he provides procedural criteria for performing EA. Firstly, there is classification according to *domain*, the breadth of context which the analyst must examine, and *extent*, the span of the utterance, or actual words, which must be changed in order to fix the error. These two dimensions of error serve to differentiate errors systematically. He also discusses the problematic area of advanced learner performance, whose language is neither fully erroneous nor expert and examines an advanced learner

corpus for error, and he suggests that 'proximate accumulation' of infelicity may cause people to perceive error in some cases.

In Bell's (2009) single-subject case study of a learner's 'wavy' form of phrasal development, the distinction between mistake and error is over-simplified and problematic, as he states that, students rarely move from error to mistake in one step: they may use a form correctly and then incorrectly, alternating several times before eventually mastering or 'knowing' a form. Another issue in relation to identification of whether an utterance contains a mistake or error is that performance is often variable, as opposed to homogenous (Ellis 1996: 48): i.e. if a student produces a correct form in some contexts, but not in others, they cannot be said to fully 'know' that form. This could be seen as an error or a mistake in a specific context or perhaps the former was correct by chance. Ellis and Barkhuisen (2005: 62) also take issue with the conclusion that a mistake of 'form' is made with omission of the -ed past simple form on one occasion with one verb, but not on another with a different verb. It is doubtful to say that they 'know the form' as the learner may have mastered the form of the second verb in the past simple, but not of the first. This raises questions of what is meant by form mastery.

Interestingly, Ellis and Barkhuisen (2005) contradict Corder's (1967) assertion that only errors are of interest to the error analyst by stating that even if a mistake has been made with, for example, the past simple form, this mistake is also important both practically and theoretically – the former because the correct form has not yet become automatised and the mistake is worthy of pedagogical attention, and the latter because 'it can be argued that the form has not been fully acquired until the learners can use it with the same degree of accuracy as native speakers' (2005: 64). This simplifies the matter for LEA purposes and therefore, errors and mistakes should be seen as one and included in the analysis.

In summary, looking at the issues in identifying errors, it may seem problematic to firstly decide if an utterance is erroneous and then differentiate between error and mistake. However, rather than abandon the potential of EA, the analyst has the option of overcoming issues related to identification by collecting written samples and ignoring awkward expression instances, omitting them from the analysis rather than over-deliberating: this will also speed up the process. Alternatively, he/she could include in the analysis all clear deviances from the standard or variety of English in which the learner is writing, regardless of whether they may

be mistakes or errors, with the rather general assumption that if the errors are written, the learner will have had more time to monitor for mistakes and they will more likely be errors than mistakes. EA can still identify the more serious errors and the vast majority of all errors, and is still therefore valuable. Dušková (1969:14) states that 'the number of cases in which it was hard to decide whether an error had been made...did not exceed 4 per cent of all the errors examined'. This figure is in sharp contrast to Lennon's findings: approximately 30% of errors made by his advanced German learners of English were considered borderline and their acceptability could not be agreed upon by the panel of speaker judges who had English as their first language. However, his errors were spoken, as opposed to written. It is my experience that, in line with Dušková's (1969) findings, only a very small percentage of errors in a written sample were considered borderline. It is important to identify all errors so that we may understand, count and categorise them. Over the last few decades various types of errors that are made by language learners have appeared in the literature, and have been categorised in different ways. This is the subject of the next section.

2.4.3 Description of errors

James (1998:96-97) offers reasons for this step in the process. Firstly, it enables labelling, naming an error as a 'type' of error provides for categorisation, and secondly, it enables quantification of errors within categories. Ellis (1997) states that accurate classification of errors assists researchers and teachers to identify language learning issues at different stages of proficiency, and to analyse different patterns of errors in different decades. This section will examine the various ways that errors have been broadly grouped since the 1960s. More specific methods of categorisation will be discussed in Section 2.4.3.1. Frameworks that were designed specifically for LEA are examined in Section 2.5.7.

The first two ways of broadly grouping errors relate to how easy to identify and reformulate they are.

Overt vs covert

Corder (1971 and 1973b:272) differentiates between two types of error: 'overtly idiosyncratic' are errors that are evidently deviant from the target language: a deviation in the form of a word or surface structure of a sentence (e.g., **I runned*: easy to spot) and 'covertly idiosyncratic', which are obviously ungrammatical only

when one considers the context. The latter is difficult to see when context and meaning are unclear. For example, **It was stopped* (difficult to spot as an error unless you know that 'it' refers to 'the wind'). Covert errors appear to be grammatical, but they do not convey the meaning that they are meant to convey. However, this need not be a great issue when analysing compositions, as context is often clear. Covert errors have meanings that are different from the intended meaning or are more related to deep sentence structure. Because covert errors can impede meaning more than overt errors, they are more serious.

Global vs local errors

Burt and Kiparsky (1974) differentiated between global and local errors. Local errors are easily spotted as their span is very short: for example, **I goed*. Global errors are less easy to identify because 'they are diffused throughout the sentence or larger unit of text that contains them, for example **We have visited London last weekend*' (James, 1998:93).

Stage of SLA development errors

Most groupings do not show how learners learn an L2. Corder's (1974) taxonomy is more interesting from this perspective. It distinguishes three error types, according to systematicity: useful for the SLA researcher, but of less use for the EL teacher:

- 1) Pre-systematic errors happen when the student is not aware of the correct form/rule. These are unexplainable by the student.
- 2) Systematic errors occur when an incorrect rule has been learned. The subject will be able to explain this erroneous rule, but will be unable to self-correct.
- 3) Post-systematic errors are those errors that happen intermittently alongside the correct form. The subject is able to explain the transgressed rule.

Whilst this list is interesting and potentially useful, it may be of limited logistical use for examining each error. The following ways of classifying error focus more on linguistic types.

Grammaticality vs acceptability vs correctness vs strangeness and infelicity

James (1998) classifies 'errors of ignorance' or 'measures of deviance' (his terms for competence errors) into four areas:

Grammaticality- This area includes errors such as subject/verb agreement (**He go to church*), and draws on rules of prescriptive grammar which help with subjectivity when deciding on whether a borderline case is an error or not. He states the grammaticality of an utterance is context free. That is, the decision concerning whether a stretch of language is erroneous or not should not depend on the context of the error. However, this is dubious, as **He go* is fine in some varieties of English, but not standard British English.

Acceptability – Whereas decisions about the grammaticality of an utterance should be context free, acceptability relies on context. James (1998:67) quotes Lyons (1968:137) when defining acceptability: 'An acceptable utterance is one that has been or might be produced by a native speaker in some appropriate context and is, or would be, accepted by other native speakers as belonging to the language in question'. James clarifies this with the example, 'Pele (the Brazilian footballer) wore a green dress'. This would be acceptable if Pele were taking part in a carnival, but unacceptable if the context referred to the kit or strip he wore when playing football.

Correctness – This refers to usage of certain 'rules' that are prescribed by some grammarians, but not always observed: for example, one should never split an infinitive, start a sentence with 'and', or finish one with a preposition.

Strangeness and infelicity – Described by Allerton (1990 in James, 1998), this category includes 'linguistically strange' word combinations, for example, **He was listening at me when I put the statement*'. They are utterances that an expert English user would usually not make unless some kind of poetic effect was being sought.

Pragmatic Acceptability

James (1998) also states we should decide on acceptability criteria (including pragmatic acceptability) as well as grammatical well-formedness. (Consider the acceptability of 'I want to read your newspaper': this utterance seems to be blunt

and would seem unacceptable to some). This lack of focus on acceptability in EA led Corder (1972:124) to state EA has attended to 'breaches of the code' and ignored 'misuse of the code.' I.e., EA can be used to identify language errors, but not etiquette errors. More recently, Thomas (1983) and Kasper and Schmitt (1996) considered pragmatic errors to be a vital indicator of communicative competence. However, decisions on grammaticality are difficult when one considers the variety of English. For example, Indian English tends to make more use of stative verbs in the continuous aspect than British English (Schubert 2002).

Incorrect vs correct by chance

A phrase may only be correct by chance, for example, in chunks, but accuracy does not transfer to constructed utterances. (Compare 'I don't know' and **I no ski*). If a learner produces both utterances, it could be concluded that they have not mastered the rules of negation. This is another issue levied at EA. However, I would argue that the lexical phrase 'I don't know' (a correct lexical phrase or collocation) had been learned satisfactorily, but that rules of negation (a grammar error) had not.

Formality

Style is another issue. Using the word 'kids' is acceptable in some written genres, but not in others. Therefore, the genre should be considered when deciding on whether an error has been made. However, this could be problematic as formality/informality is not polar and some analysts may have different ideas of acceptability in terms of style.

Formal vs semantic

Formal errors are errors in grammar, pronunciation or spelling: semantic errors are errors in meaning, such as using an incorrect near synonym. This seems to be a useful dimension of categorisation, as the two could be quite mutually exclusive. Various researchers (e.g., Hemchua and Schmitt, 2006) have used this framework.

There are further linguistic methods of classifying types of errors:

- The skill or modality (proficiency in speaking, writing, reading, listening).

- Linguistic levels or language system
(pronunciation, grammar/syntax, vocabulary, style or discourse).
- Cause (interference, interlanguage, induced) (see Section 2.4.5 below).

2.4.3.1 Frameworks for error categorisation

This section will examine some of the main ways that whole language EA classification taxonomies (as opposed to purely LEA) have been produced. LEA taxonomies are dealt with in Section 2.5.7.

Corder (1973) categorised errors into four areas: (1) deletion or omission of some required element that should be present; (2) addition of some unnecessary items; (3) selection of a deviant item and (4) mis-ordering of items.

Category	Description	Example
Omission	Something missing	<i>She *sleeping</i>
Addition	Something added	<i>We didn't go *there home.</i>
Misselection	The use of the wrong form of the morpheme or structure	<i>The dog *ated the chicken</i>
Misordering	The incorrect placement of a morpheme or group of morphemes in an utterance.	<i>What *daddy is doing?</i>

Table 1 Corder's (1973:278) Surface Error Taxonomy

James (1998) added a fifth category, blend, to this taxonomy. A blend is where two items are blended to produce an error (*..the *depths of the ocean*. Blending deep and depths). This is a surface strategy taxonomy, which 'highlights the way surface structures are altered' (Dulay *et al*, 1982:150). Dulay *et al* state that this is 'promising because it provides an indication of the cognitive processes that underlie the learner's reconstruction of the L2'. However, this is also problematic, as it suggests that 'learners operate on the surface structures of the target language rather than create their own unique structures' (*ibid*). Although there have been many attempts to perform EA using this system, unfortunately, this does not provide much depth of linguistic analysis: the teacher or analyst would want to know of sub-types of these categories. Corder himself states that such a classification is superficial, and insufficient to describe errors. Even with James' (1998) suggestion for a fifth, blends, this taxonomy still seems rather minimalist, as there is still too little that the teacher or researcher could use to describe the types of errors in more detail and inform understanding of language learning or develop teaching strategies. For information of this type to be of use, these categories

would have to be further sub-divided: a linguistic taxonomy is required. Corder's taxonomy would not point to areas that a learner could be asked to work on, such as prepositional partners or subject/verb agreement. It is the details of what is omitted that teachers and SLA researchers are interested in.

Dulay *et al.*, (1982) created a taxonomy comprising four groups of errors: (1) linguistic category (see below for examples); (2) surface category (Corder's taxonomy); (3) comparative analysis; (4) communicative effect. For pedagogical and SLA research, perhaps the first, linguistic category could be of greater use, but one could envisage dual categorisation issues with this framework, as there would inevitably be errors that fall into two or more of these areas. Whereas a linguistic category would be the most useful, there are several different models to choose from.

A basic linguistic taxonomy involves lexical and grammatical errors (see Table 2 below for an example). Corder (1971) states that there should be a linguistic level in which sub-categories such as morphology, syntax, and lexicon are included. Politzer and Ramirez (1973) distinguished between errors of morphology, syntax and vocabulary and found that LEs exceeded grammatical ones. The errors that Politzer and Ramirez's found were then placed into sub-categories according to parts of speech (p41). Meara (1984) also found three times as many LEs than grammatical ones.

Burt & Kiparsky (1972) created a six main category taxonomy: (1) skeleton of English (missing and mis-ordered parts); (2) auxiliary system; (3) passive sentences; (4) temporal conjunctions; (5) sentential complements and (6) psychological predicates. (e.g., Misordering with reverse psychological verbs **The cat is on the dinner table but my father doesn't bother *that.*) Sub-categories were made, e.g., aux verbs are put into 'do' have' and 'be', modals, etc. This type of taxonomy allows for detailed description and quantification and serves the teacher planning remedial instruction well. However, a more comprehensive taxonomy may make categorisation difficult as it may allow for dual categorisation of error. Furthermore, the newly qualified teacher may struggle with this level of grammatical metalanguage.

Level modification	Substance Phonology Graphology Spelling Punctuation	Text		Discourse Cohesion coherence genre-fidelity felicity
		Grammar Rank: clause- phrase-word- morpheme. Class: noun, verb, adjective, adverb, preposition, conjunction, etc.	Lexis Sense relations collocations	
Omission				
Over-inclusion				
Misselection				
Misorder				
Blend				

Table 2 James' (1998:274) Framework for Error Analysis (target modification taxonomy)

James (1998:114), drawing from work done in the field of language therapy, also suggests combining linguistic categories and surface structure categories (or target modification taxonomy, as he calls it) to produce a method for EA that is more comprehensive and produces a linguistic profile on the subject at hand. (See Table 2). This method seems superior to other suggestions, due to its comprehensive and clear nature, but it also seems rather time-consuming for the busy EFL teacher or SLA researcher to use this as a taxonomy when dealing with, for example, 20 or so texts of approximately 350 words each.

Corder (1981) states that the more appropriate the linguistic taxonomy, the better the linguistic description of errors will be. Certainly, appropriacy of category is the key, but this could be quite difficult to achieve. However, it could also be that the more comprehensive the taxonomy, the more difficult it would be to accurately allocate all errors as they may be classifiable into more than one area. A balance must be struck between how manageable/large the taxonomy is and how useful or detailed the results are for researchers and teachers.

The process of classification

Corder (1973:274-275) states that one needs to reconstruct each error before a description can be made. To avoid ambiguity on the part of the analyst, he suggests an authoritative reconstruction by asking the learner to state what he/she

meant in his/her mother tongue. This is then translated into the L2 to provide a correct form. Or in the absence of the learner, the analyst could supply a plausible reconstruction, based on the error, the co-text and the learner's background, the learner's state of interlanguage, and his understanding of the world and the L2. If neither reconstruction is possible or satisfactory, this data has to be excluded from the analysis. However, this is unsatisfactory if one takes the view that the errors that are incomprehensible are the more important as they cause breakdowns in communication. Therefore, a miscellaneous or phrase error category should be included so that the error is not overlooked in the analysis, but can be dealt with at a later stage. Corder's formal stages in EA may be quite lengthy, so the practicing teacher may, in the absence of the learner, wish to pursue a short cut, and reconstruct erroneous clauses mentally as they analyse the data: a plausible reconstruction. However, the formal recording of the reconstruction does allow for more careful analysis.

Ellis (2008:50) states, 'In order to accurately identify the type of error made, one must compare the learner's idiosyncratic utterances with a reconstruction of those utterances in the target language or, more recently, with a baseline corpus of native-speaker language'. However, writing out the correct form is not always practical, especially when the analyst is a busy, practicing EFL teacher.

Furthermore, overt error identification is problematic **I am worried in my mind* = I feel worried or I have a problem on my mind? How this error is classified will depend on how it is reconstructed. As mentioned above, even asking the learner can be problematic. It is doubtful that when asked, they would be able to come up with either of these two alternative meanings of the error above because if they were aware of these meanings of these two phrases, they would have used one of them initially. Covert error reconstruction is even more difficult. Perhaps the exact wording is not so important at this stage. The intended meaning in this case is clear enough. One of the alternatives above could be presented to the learner and he/she could be asked to choose. The issue here is one of categorisation and the error appears to be one of collocation or phrase, which calls for a category titled as such. It would make sense to categorise the error, rather than the reconstruction.

Another issue relates to reliability, or testing of categorisation of frameworks by asking other raters to perform EA on the same scripts and comparing results.

There have been very few of these: perhaps because they are of less interest to

editors of publishing journals. Problems of study comparison are further compounded by the fact that many analysts create a new taxonomy of categories when they perform EA. This is the most problematic and inconsistent area of EA: in some frameworks, errors could be attributable to more than one category, while in other frameworks, there are insufficient categories for categorisation of all errors.

This section has examined some options for whole language EA and raised further issues with EA. The next section deals with counting or quantification of errors.

2.4.4 Quantification of errors

Some studies have omitted quantification of different types of errors (Jain in Richards 1974 and Richards 1971). Dušková (1969) only offers counts for main categories, and not subcategories. Schachter and Celce-Murcia (1977) state that for these numbers to be useful, we need to know how many errors learners could have made in these categories – in other words we need relative, rather than absolute error frequencies. This is also problematic as it is doubtful that two different raters will agree on the final number, given the issues raised in Section 2.4.2. It is also an issue as there are many ways of making errors with a single sentence. One suggestion from Ellis (1996) for an 'error score' is to produce a ratio of erroneous clauses to error-free clauses. Some studies of EA (Hemchua and Schmitt 2006, for example) count error types rather than tokens. These SLA researchers are interested in what kinds of errors are being made by a certain cohort of learners, but James (1998:114) feels it would be 'invaluable' to add the tokens. This would certainly make EA of much more use to the teacher and learner, as they would then be able to see how many of certain types of errors were being made. This information could inform remedial teaching to a great degree. Lennon (1991) suggests avoiding repeat counts if the error is a lexical replica. He makes no suggestions for counting grammatical types or tokens.

Counting error-free clauses to produce an accuracy score may appear out of line with the approach taken by Hemchua and Schmitt (2006), who counted error-free words. However, is more in line with a view of lexicogrammar because some lexical errors span across a phrase or occur as multi-word units, such as idiomatic expression errors, making comparison of total number of errors with total number of words problematic. Furthermore, it is easier to identify clause breaks in a sentence and errors within clause breaks than counting individual errors.

2.4.5 Explanation of errors

This stage involves establishing the source of the error, or saying 'why' it happened – another difficult, but important stage for SLA researchers, as it attempts to establish the processes responsible for L2 acquisition.

Main causes of errors

A survey of the literature shows that there are several causes of errors. The main three appear to be: either transfer (based on L1 interference); intralingual or analogical errors, caused by ignorance of a target language rule or lexeme, or inability to implement that language; and finally, induced errors, which are caused by inadequacies in the teaching and learning situation (Richards 1971b; Corder 1974 and Dulay and Burt 1974b). These three areas will be examined before looking at alternative ways that authors have grouped errors by causality.

Transfer errors

Transfer or interlingual error refers to errors where the learner's output is influenced by the effect of her first language (L1). When she is unaware of the correct form in the target language (L2), she may introduce syntax from her L1 or use a borrowed phrase, coin a phrase or translate directly (calque). These are sometimes more positively known as code-switching or communication strategies. According to the Contrastive Analysis Hypothesis, structures that are similar in the L1 and L2 will be less difficult to acquire than those which are very dissimilar. The first type will result in positive transfer and the second in negative transfer. These types of errors occur at all proficiency levels and may manifest as phonological, morphological, grammatical, or semantic errors (Richards 1971). However, it must be noted here that not all transfer is negative. The next section discusses positive transfer.

Cross Linguistic transfer/influence

Following Odlin (1989: 27) I define transfer here as 'the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired'.

For several decades, beginning with Fries' (1945) argument for contrastive analyses, the study of how the learners' first languages has been used to not only

inform understanding of the errors that they are likely to make, but also the positive influence that it may have. For example, 'like English, Swedish has an article system whereas Finnish does not, and so Swedish speakers should, if the prediction holds true, tend to do better [in learning about and using English articles]' (Odlin 2022: 5). This process of analysis is known as cross-linguistic influence or transfer. It is not limited to lexical items (often false-cognates), but could also manifest as phonology, syntax, pragmatics or morphology (Benson 2002). The very useful *Learner English, a guide to interference from other languages*, now in its second edition, has provided insight in many of these areas for practicing teachers for several years.

Language Transfer is highly complex and research into users of several languages has shed new and recent light onto this area (e.g. De Angelis & Dewaele, 2011; Gabryś-Barker, 2012). It overlaps with and can be confused with codeswitching (Odlin 2022). However, the basic points, according to Gass and Selinker (1992), are as follows: where two languages share similarities, there can be facilitative, positive transfer. However, when a structure in the target language does not exist in the learners' first language, learners may avoid that structure. More interestingly, these differences can lead to different rates of development, i.e. delay of development in comparison to other L1 groups when the learners' L1 has certain different structures to English or acceleration compared to other L1 groups when structures are more similar between the two languages. It can also lead to different routes of acquisition where learners whose first language does not have certain structures may use different structures in non-standard ways. Finally, it can also lead to overproduction of near synonyms when they exist in both languages.

The frequency of cross linguistic transfer seems to depend on several factors, such as setting (more common in classrooms due to paucity of input and interaction opportunities when compared to naturalistic settings), learner proficiency (mainly more prevalent at lower levels), style (oddly more prevalent in a more careful, monitored style than in spontaneous interactions and learner type (learners who are risk-takers, meaning oriented and those with a positive attitude to the L2 are less prone than form focussed or those with a negative attitude). Transfer can occur consciously as a communication strategy or unconsciously through a lack of awareness or learning of the correct form. An interesting cause, apart from interlanguage (proficiency is still developing) and the influence of

existing language knowledge is loss of learner identity if the language is learned too well.

Because of its association with an outdated view of language acquisition (behaviourism), language transfer fell out of fashion, but according to Gass and Selinker (1992), 'There is now overwhelming evidence that language transfer is indeed a real and central phenomenon that must be considered in any full account of the second language acquisition process'.

Intralingual errors

The second type is broadly agreed upon as being analogical or intralingual errors: these errors are caused by lack of awareness or failure to implement a rule of form. These could be performance or competence errors. They are caused by incorrect or partial learning of a feature of the L2. This type of error is exceedingly common in many proficiency levels and in learners from all L1 groups (Richards 1971). According to James (1998), they are due to learners resorting to a strategy or trial and error system to help them to communicate and understand L2 rules and structures, but they are often causes of errors. James lists the following types of intralingual errors:

- 1) false analogy (assuming a new item behaves like an old one)
- 2) misanalysis (setting false hypothesis due to misinterpretation of the target language)
- 3) incomplete rule application (or under-generalisation)
- 4) exploiting redundancy
- 5) overlooking co-occurrence restrictions
- 6) hypercorrection (consistent over-monitoring L2 output)
- 7) overgeneralisation or system simplification (overuse of some specific items/rules and underuse of other similar items).

Induced errors

Schuman & Stenson (1974 in Spolsky, 1979) first wrote about exigencies, or induced errors, which are caused by the teaching and learning situation rather than difficulty of L2 or transfer from L1. Schumann and Stenson state that although these errors may not be able to inform the SLA researcher about language acquisition, they are important as they often slip by unnoticed and may hide a

learner's true ability to learn a language. Left untreated, they may fossilise, or become entrenched and very difficult for the learner to self-correct (p126). James (1998) identifies five types of induced errors:

- 1) materials-induced errors.
- 2) teacher-talk induced errors.
- 3) exercise-based induced errors.
- 4) errors induced by pedagogical priorities.
- 5) lookup errors where learners have misused reference materials, such as dictionaries.

Allocating causality is often problematic as the process is time consuming and subjective. If one were proficiently bilingual and working with a single L1 group, then one may be able to allocate causality to some of the interference (e.g., calque, borrowings or coinage) errors, or one could interview learners about how they arrived at the error. Other than that, allocating to cause would be quite subjective in all other contexts or situations. Furthermore, although several authors (Corder, 1967, James, 1998) have attached much importance to identifying causality, it is not clear how this information will help learners in the classroom. One might also question the value of attributing cause to errors when providing feedback. For example, telling learners they have made an error because of their first language will not necessarily help the learner overcome the error. The learners themselves may very well understand their own thought processes that led them to the error. Therefore, this stage could be omitted from EA, as it is rather difficult and inaccurate and of outdated interest. Omission will also make the process more streamlined.

2.4.6 Evaluation of errors

According to Ellis (2008), this stage involves examining the effect(s) that the error has on participants: how well they have understood the meaning or what their affective response is in terms of how the error makes them feel (confused, irritated, for example). Evaluation studies were quite common in the 70s and 80s, and were motivated by improvement to language pedagogy. Ludwig (1982) offers a detailed description of such studies. Error evaluation studies include participants, such as expert and developing English users, who 'judge' the errors in different ways. Different error types have been evaluated (semantic, lexical, grammatical,

spelling) and judgement criteria have also varied (comprehensibility, seriousness, naturalness or degree of irritation: Santos 1988). Sometimes judges are asked to correct the errors and give reasons why they judged some errors as especially problematic. Sometimes judges' comprehension of errors is tested. Instruments used to elicit judgements vary (decontextualised lists of errors, contextualised language, presented orally or in writing).

According to Ellis (2008), error evaluation studies often ask:

- 1) Are some errors more problematic than others?
- 2) Are there differences between NS and NNS evaluations?
- 3) What criteria do judges use in evaluating learner errors?

Johansson (1973) stated that errors should first be evaluated by whether they are comprehensible and second, whether they cause irritation. Secondary factors include frequency and generality. Johansson was able to then create a hierarchy of errors but stated that 'it is not possible to illustrate the scale of errors at the present time since there is no available information concerning the degree of comprehensibility /irritation caused by different errors' (1973:109). Findings for error evaluation studies show that:

- a) Expert language user judges say LEs are more serious than grammatical ones (Burt, 1975; Tomiyana, 1980 and Khalil, 1985) and global grammar errors (affecting overall sentence organisation, e.g., word order, missing or misplaced sentence connectors and syntactic overgeneralisations) are more likely to interfere with comprehension than local ones (affecting single elements in a sentence, e.g., morphology). As Santos (1988) reports, however, there have been conflicting opinions on this.
- b) Those who do not have English as their L1 are much more critical than those that do (James, 1977), particularly on morphological and function word errors, but those with English as L2 evaluate lexical and global errors less seriously than L1 English language users.
- c) Judges use different criteria in evaluating gravity of error. Khalil (1985) identified three general criteria (intelligibility, acceptability and irritation). English L1 users are more concerned with the effect of error on comprehension. Others are more influenced by transgression of what they understand to be the rules of the language.

The general conclusion of error evaluation studies is that teachers should, quite rightly, pay more attention to errors that impede communication (i.e. semantic and global grammatical errors). However, lack of contextual descriptions when asking judges to judge error seriousness leads to a lack of reliability, and it is still unclear what criteria individual judges use when asked to assess 'seriousness', 'intelligibility' or 'acceptability'. Teachers performing class EAs will form their own opinions on error seriousness. These opinions will surely be formed by the teachers' beliefs in terms of whether they believe surface or global, or overt or covert errors, for example, to be more serious. Seriousness may also be determined by the perceived learners' goals (e.g., pass a grammar test or achieve a suitable level of communicative competence to work in a specific role). Therefore, even if a hierarchy of error seriousness could be agreed upon, it may not be suitable for all contexts of language teaching and learning. Therefore, this stage of EA could be omitted from the search for a universally applicable framework for LEA and decisions on seriousness should be left to the teacher who is better placed to understand the learners' needs and context.

2.4.7 Summary of the limitations of EA

During the 1980s, EA fell from popularity due to the procedural issues discussed in Section 2.4). The main critics of EA, according to Ellis (2008), are Bell (1974), Schachter and Celce-Murcia (1977), Long and Sato (1984) and Van Els *et al.* (1984). Below I summarise the main issues and offer a defence of EA.

Variety

EA takes some target language variety as its reference point, which may be problematic if, for example, Thai speakers, living and working in Thailand wish to speak 'Tinglish' (Thai English), but their output is judged against 'correct' British English. Furthermore, analysts will not always know what variety of English the learner is using. Bley-Vroman (1983) stated that a learner's output should be thought of as an idiosyncratic dialect, echoing the thoughts of Corder (1973), 'a variety in its own right' and 'EA is guilty of the comparative fallacy: that is, it seeks to account for learner language in terms of target-language norms'. However, this seems like a rather pedantic point, as this could be applied to all assessments of learner output. Surely it would be sensible to compare the output of learners of Hispanic English with Hispanic English itself and it would make sense for students studying English in UK to compare their interlanguage with an approximation of

British English. James (1998) also counters this point by stating that 'learners are typically targeted on native-speaker norms and as such, perform 'cognitive comparisons' in the process of learning an L2'. In response to Bley-Vroman (1983), language teachers and learners need models for language teaching and learning. Jenkins (2006) has suggested a standard variety of World English. This seems to be a suitable model for learning and teaching. However, if EA were to be performed in a country that had its own variety, such as Indian English, it would make sense if that variety would provide the standard for EA, not American nor British English, for example.

Incompleteness

'EA fails to provide a complete picture of learner language. We also need to know what they can do correctly' (Hammarberg 1974:185). EA is only able to analyse output, not receptive skills. It cannot shed light on errors of pronunciation when writing is being analysed. I believe that it would be odd to criticise a test of written English, for example, for its lack of focus on pronunciation ability. Hammarberg (1973:34) states that 'EA can have a place as a partial and preliminary source of information at an initial stage of investigation'. Corder himself (1974) states that EA is not intended to be a comprehensive examination of proficiency, but is designed solely for the purpose of analysing deviations from the norm in learner output. However, the teacher or researcher, if taking a quantitative approach to EA, could count the number of error free clauses or provide qualitative comments on what was done well, as is done with other evaluative rubrics. In the absence of a suitable alternative to EA, it makes sense to try to ameliorate issues with the existing system. A suggested alternative to EA, Obligatory Occasion Analysis only targets specific target language structures, and takes no account of when a learner uses a target language feature in a context that is not obligatory (Ellis, 2008), so it seems logical to pursue a working EA framework.

Limitations in scope

Most studies, such as this one, are cross-sectional, offering a static view of language acquisition. This does not help us understand how learners develop over time. Many studies have been cross-sectional, a snapshot in time of the learners' current interlanguage development, making it impossible to see the different errors different learners make at different stages of their development (Ellis, 2008). There are indeed very few longitudinal studies. Lach's (2011) is discussed in Section

2.5.7.12, but in Chamot's study (1979), the findings were inconclusive, which underlines the need for more EA studies and should not be seen as a criticism of these studies.

Problems in error identification

According to Norrish (1983:91), error identification in terms of acceptability is quite subjective. Furthermore, language, or more specifically lexis can 'come into fashion', making previously unacceptable language acceptable later when it is included in mainstream dictionaries (e.g., the meaning and use of the term, 'Awesome' has evolved). Norrish also states that the rules relating to certain features of language, e.g., use of commas are ill-defined, making error identification subjective. According to Shachter (1974), it does not analyse use of communicative strategies, e.g., avoidance of 'difficult' forms. Error identification can be problematic, especially in a spoken corpus with regard to global errors (Lennon, 1991). De Rocha (1980 in James, 1998) suggested that to avoid subjectivity, putative errors could be submitted to a panel of judges for consideration. Norrish (1983:92), on the topic of acceptability, states that simply because an English as a first language user would not utter a phrase, does not mean it should be dismissed as an error, especially if the meaning is clear, it is grammatical and does not cause irritation. This seems to be an acceptable set of criteria to help with borderline cases. However, Taylor suggested that 'what constitutes significant error is not strictly quantifiable' (1986:162). This is not a new criticism of EA. Whilst it is true that a quantitative only approach of counting errors does not take into account issues, such as organisation and attempt at complex grammatical constructs at a level above the learner's proficiency level, it is not claimed here that EA should be the only form of analysis and feedback on learner writing. Lennon prefers the quantification of errors, showing how some problems with identification can be overcome. He points out that most 'erroneous forms are, in fact, in themselves, not erroneous at all, but become erroneous only in the context of the larger linguistic unit in which they occur' (1991; 189). As stated in Section 2.4.2, Lennon proposes two new dimensions of error: domain (the breadth of the context, including word, phrase, clause, previous sentence or discourse which needs to be considered for determining whether an error has occurred) and extent (the size of the unit: morpheme, word, phrase, clause or sentence that needs to be deleted, supplied or reorganised in order to repair an error). Dušková (1969), as previously mentioned, acknowledges this issue by stating that 'the

number of cases in which it was hard to decide whether an error had been made...did not exceed 4% of all the errors examined'. Error identification is an important and central, yet problematic stage for any EA. It would be interesting to establish whether, with detailed guidance, raters could agree on what language is erroneous. Overcoming issues of identification is the subject of much of this dissertation.

Problems of error categorisation or typologies

From looking at the error itself, it may be difficult to establish the type, as some errors can be allocated to more than one category. Furthermore, researchers conducting EA often create a new typology of their own, making comparison between studies quite difficult. Corder (1973) suggested questioning the learner as to her intended meaning, but this would be time-consuming and is not always possible, and may prove to be unreliable if the learner cannot remember the intended meaning. The solution here may be to use context to infer probable meaning, to create a plausible reconstruction (Corder, 1973). If this does not clarify meaning, the error could be allocated to a new category (incoherence) or rejected from the analysis. However, rejection would be unsatisfactory, as this would eliminate the most serious of error types, those which cause breakdown in coherence, from the analysis. More replication studies that carefully use the methodology of the original study (see Study 1, next chapter) are required. This way, the reliability of an EA framework could also be tested. Categorisation also remains a central but problematic stage in EA. Attempts to overcome related issues are also the focus of much of this dissertation.

Despite the fact that EA is still in use, for the reasons above, it has largely been abandoned since the mid to late 1970s. In the mid-1970s, Interlanguage established itself as a more popular theory or approach to language analysis. Two further methods or replacements for EA have been put forward: Obligatory Occasion Analysis, which examined what the learner could do well, as well as the errors they were making (Ellis and Barkhuizen, 2005).

Obligatory Occasion Analysis (OOA)

According to Ellis (1996: 716), OOA 'involves identifying contexts that require the obligatory use of a specific grammatical feature in samples of learner language and calculating the accuracy with which the feature is actually supplied in these contexts'. A comparison is made between how the learner uses a specific linguistic

feature and how it is used by expert users of the language (Ellis & Barkhuizen, 2005: 73). This addresses one of the issues with traditional error analysis in that it acknowledges what the learner does well as well as where required language has been omitted. It has also often been used in L2 morpheme studies for establishing the natural order in which learners universally acquire L2 forms. Because the procedure focuses on a finite set of structural or morphological features of language, it is of limited value when analysing lexical errors, unless one is researching the use of one specific lexical item. However, the scoring aspect of the process is of interest and could be adapted for LEA: a text is analysed to ascertain the number of times a target item was required. A score is calculated by dividing the learner's correct uses by the total amount of obligatory occasions and expressed as an accuracy percentage, which will provide quantitative feedback. This approach could be used to calculate clauses and lexical error-free clauses in a piece of writing, again, providing a score. Because OOC does not account for instances where a form was overused by a learner, another method was created by Pica (1984 in Ellis and Barkhuisen 2005). This was the target-like use analysis method where the amount of the learner's correct uses of the target feature is divided by the number of times it should have been used plus the amount of times it was overused. Furthermore, obtaining an idea of how well used a specific feature of language is will also help the analyst to decide whether misuse of that feature constitutes an error or a mistake, as discussed in Section 2.4.2.

Some authors, including Al-Khresheh (2016:49) feel that 'despite the criticism that EA has received, it still plays a fundamental role in investigating, identifying and describing second language learners' errors and their causes'. Furthermore, despite the perceived failings and demise of EA, there has been a rising interest in the process since Lennon suggested revisions to the process in 1991, i.e., Leki (1991), Selinker (1992), Teh, (1993), Kellerman (1995), Brown (2000) and Jiang (2009).

This section has discussed some of the criticisms levelled at EA. Some of these criticisms, such as issues with identification are completely valid, yet others, such as incompleteness seem harsh, and when taken together may have caused the abandonment of a potentially useful analytical tool, when what could be beneficial is to re-evaluate its not inconsequential contributions to teaching and learning,

salvage the working components of EA, attempt again to overcome some of the more serious issues, and reconsider how it may be best employed in this area.

LEA has attracted some interest in recent years (Yang and Xu, 2001; Hemchua and Schmitt, 2006; Llach, 2011 and Picot, 2017), but no significant attempt has yet been made to produce a reliable framework that overcomes the more serious issues above.

2.5. Lexical error analysis

This section will provide a brief rationale for a focus on lexis and LE, and will examine the main frameworks for analysing lexical error that have been published since the late 1960s. The frameworks are described and examples of errors in the various categories are provided.

2.5.1 The importance of lexis

Wilkins states that 'Without grammar very little can be conveyed, without vocabulary nothing can be conveyed' (1972:111). Wilkins ignores the communicative contribution of gesture, but this quote has often been used to underline the central role that vocabulary has in communication. Before the 1990s, lexis was considered the 'Cinderella' of language teaching (Beheydt 1987), and in language education much more emphasis was placed on grammar instruction and testing. Meara (1980) has a similar view. It was thought that if learners could grasp the fundamentals of grammar, they could pick up the vocabulary as their learning progressed. However, since then the teaching and learning of lexis has taken a more central position, and there has been an expansion of research into vocabulary acquisition (Kamiya and Nakata, 2021:387). This is largely due to progress in corpus linguistics and the growing realisation that language frequently appears in the form of lexical chunks which play an important function in communication (Sinclair, 1991; Biber *et al.*, 1999 and Hill, 2000). This central view of lexis is also partially due to the effect of Michael Lewis' ground-breaking book, *A Lexical Approach to Language Teaching* (1993). It was Lewis who coined the phrase 'Language is made up grammaticised lexis, not lexicalised grammar' (p34). Lewis (1993:89) also states that, 'Lexis is the core or heart of language but has always been the Cinderella'. By this, he agrees that the teaching of lexis has been ignored in some of the main approaches and methods of language instruction over the years (e.g., Grammar translation and audiolingualism). It can still be seen in

many modern coursebooks that grammar is the main organising principle in syllabus design.

2.5.2 Lexis and grammar

Perhaps one of the most important reasons for a focus on LEA and a shift away from a main grammar focus in EA is the changing view of language itself. The traditional boundaries between grammar and lexis have become blurred. This view was popularized by Michael Lewis (2000:137) with his previously stated and famous quote 'Language consists of grammaticised lexis, not lexicalised grammar'. For instance, it would be futile to attempt to analyse the grammatical errors in the phrase, '**It rain with the cats and dogs*', as this is a chunk of language or a single lexical item, which should be learned as an unanalysed whole (Schmitt 2004; Wray, 2008; and Thornbury, 2019). The same principle could be extended to the full range of lexical items in Lewis' 1992 taxonomy for other errors, for example, ones that could be classified as spelling and L1 influence. Grammar and lexis are completely interdependent. This view is supported by computer linguistic analysis (e.g., Hunston *et al.*, 1997), who suggest 'bridging the gap between lexis and grammar' and talking instead of 'language patterns'. They stated that lexical items 'should be described in terms of patterns, and that words sharing patterns often share meanings too'. Lewis (1993:142) also stated that 'every word has its own grammar – a set of patterns in which it occurs'. Salem (2007:212-213) cites Lewis and writes that 'errors differ in their degree of 'word-sensitivity''. This term refers to the extent of generalisability of a rule that has been infringed. Word-sensitive errors are those that can be attributed to infringement of a word-intrinsic requirement. For instance, **enjoy to speak* is a word-sensitive error; it can be attributed to violation of a word-intrinsic restriction of 'enjoy' not to be followed by an infinitive'. Therefore, word grammar can be analysed with LEA. Lewis states that corpus linguistic analysis has helped us to understand that language is not organised mainly by grammar, but by different types of lexical chunks of language (Lewis, 1993). For example, the following examples of collocations could be considered to be grammaticised lexis as collocations often consist of:

- Adjective + noun (stale bread, tight security, vital evidence)
- Noun + noun (pressure group, death threat, chicken curry)
- Verb + adjective + noun (play a significant role, report a significant fall)
- Verb + adverb (live dangerously, work hard, sleep soundly)

- Adverb + verb (seriously affect, deeply offend)
- Adverb + adjective (highly successful, well established, fully informed)
- Verb + preposition + noun (go on display, be on the move)

This analysis clearly supports a concept of word grammar and therefore a focus on LEA. It also gave rise to the concept of 'lexicogrammar'.

2.5.3 Lexicogrammar

Richards (2015: 301) also states that the 'boundary between them [lexis and grammar] is not rigid' and in order for learners to use the lexis that they have learned accurately, they need to know not only their part of speech, but also what grammatical patterns words fit into, such as 'which verbs can and cannot be used with the present continuous tense, which are transitive or intransitive, which nouns are countable or uncountable, which adjectives can be used predicatively and which cannot, and what the normal order of adjectives is' (2015:301). 'Learners also have to learn the many different affixes that change the grammatical function of words. Knowing the grammatical function of these affixes will help learners to identify the meanings of new words' (Richards, 2015:302). Liu, and Nelson (2016) write convincingly about the importance of teaching language as an integrated system of lexicogrammar. They state that grammar and lexis are positioned at either end of a continuum and should not be considered as entirely separate entities. However, in many cases, they clearly are. Hunston and Francis, (2000) state that 'syntactic structures are often lexically confined, while the use of a lexeme almost always has grammatical implications'. They offer the following example to underline this point: '*enjoy* and *love* are near-synonymous verbs, but while *love* can take as its object either an infinitive or a gerund (love to read books and love reading books), *enjoy* may take only a gerund object, enjoy reading books'. A large number of studies of corpora have found a very close interconnection between the two areas (e.g. Biber *et al.*, 1999 and Hunston and Francis, 2000), which strongly acknowledges a systemic-functional linguistics perception of what we say or write as a system of co-constraining selections affected by grammar and lexis simultaneously. Cognitive linguists see our speech as a symbolic system made up of symbolic constructions. Liu and Nelson (2016) state that grammar and lexis should be taught together with a strong focus on multiword units and constructions, as well as the 'structural patterns within which a lexical item typically occurs'. This view is supported by Howarth (1998) and Lewis, (2000). Truscott (1996:34) states that lexicons consist of 'much more than lists of

words: the form, meaning and use of each word depends very much on its relationship to other words and to portions of the language system, as well as to non-linguistic cognitive systems'. The rise of grammaticised lexis and the importance of collocation in teaching is also supported by James (1998:142-143) and many others. If these views of language are accurate, it follows that error analysis should focus on lexis instead of (or at least as well as) grammar.

2.5.4 Lexical error analysis vs whole language error analysis

Interestingly, Candler (1979:267) points out a problem associated with general or whole language EA, as opposed to LEA. He states that if a teacher devises remedial teaching based on the findings of such investigations, particularly on the frequency or most common error types found, she is in danger of focusing on the least serious errors, perhaps structural errors, such as subject verb agreement and not on the type of errors that affect meaning, which may be less frequent and difficult to categorise. He cites Widdowson (1972) as stating that teachers should devote more attention to the value of communicative acts and less to *signification* or the structural aspects of language. Because lexis carries more (and some different) meaning than grammar, it makes sense to focus our attention more on the lexical, rather than the grammatical aspects of the language.

2.5.5 A justification for LEA

Kamiya and Nakata (2021:387) report a general lack of studies into how corrective feedback on learner compositions can aid acquisition. It is hoped that LEA will facilitate future studies. However, not all researchers see value in supplying corrective feedback on learners' LEs. This claim appears to me to be abandonment of teacher duties and whilst it is true in some situations, may not be applicable in all. Kamiya and Nakata (2021:394) state that 'written corrective feedback may lead to appropriate use of vocabulary in revisions as well as rewriting on the same prompt', and according to Swain and Lapkin (1995), Storch and Wigglesworth (2010) and Coyle and Roca de Larios (2015), learners do notice written corrective feedback on LEs in written compositions. There are clear benefits of providing immediate and delayed written corrective feedback (Hanaoka, 2007). Sheen (2010) argues for direct, explicit, rather than implicit,

feedback. LEA would provide such direct and explicit feedback and it would be interesting to measure the vocabulary learning gains when tested.

LEA can inform classroom teaching. For example, Hemchua and Schmitt (2006) were able to identify the types and possible causes of their Thai students' LEs. This resulted in recommendations for teaching. Lee (2017) used pairs of commonly confused words, such as wound/injure (from Hancock, 1990) to help her Korean learners efficiently identify and avoid these common errors through the use of pictures. Without research into LEs, this innovative and very useful technique would not have been possible. Llach (2011) sought to find a correlation between LE and essay score (young Spanish learners' writing). She hoped to provide objective criteria for evaluation and information on what to focus classroom instruction on. Findings would show learners what to pay more attention to. Her work attempted to use LEA to measure other areas of results of teaching, such as ability to write and knowledge of receptive vocabulary.

James (1998:143) offers five further sound reasons why LEs are of particular significance and why LEA should be undertaken:

- 1) Lexis is taking a more central role in language study (morphological aspects of words, multi-word lexical units including, idioms and their centrality to expert fluency).
- 2) Language learners themselves place great importance on learning the lexical system of a language, perhaps more so than grammar and pronunciation.
- 3) LEs have been proved to account for the greatest percentage of errors in student output. Llach also states that LEs are found to be the most numerous and most serious in many different research studies (2011: xi-xiii). This is supported by Grauberg (1971), Meara (1984) and Lennon (1991 in Hemchua and Scmitt (2006). This point may be debateable as it appears that the type of error made varies with proficiency level and certainly with the framework used to establish types of errors made.
- 4) Those who have English as their L1 consider LEs to be the most disruptive and irritating of all error types. LEs are the most 'irritating' (Santos, 1988 in Hemchua and Schmitt, 2006) and are less tolerated than syntax errors (as LEs contain content words) outside the classroom (Carter, 1998:185 in Hemchua and Schmitt, 2006).

- 5) Lexis carries a much greater communicative function than other language systems. Whilst this may be a moot point, it is certainly my opinion, also.

Llach (2011:xi-xiii) offers further reasons: 'lexical errors can inform us of a learner's writing ability, quality predictor and lexical knowledge'. Therefore, we need a system that is more reliable and objective that we can use to assess writing. Therefore, we need a more systematic method of examining LEs. (Llach, 2011:xi-xiii). Llach also states that 'vocabulary's functional role in communication and assessment is crucial' (Llach, 2011: xi-xiii) and that there is a lack of research into LEs in recent years. Longitudinal studies are almost non-existent (Llach, 2011: xi-xiii), making this an under-researched area, worthy of investigation. Hemchua and Schmitt also support this 'there has been little research in LEA in L2 writing' (Hemchua and Schmitt, 2006). We require more studies that look at the role of EA in writing assessment, lexical knowledge and L2 vocabulary acquisition (Llach 2011: xi-xiii). 'Learners can learn from their errors by spotting problematic L2 areas where they need more practice' (Llach, 2011: xi-xiii). Other authors (Laufer, 1991:321 in Llach, 2011:xii) also support more investigation into LEs, as they 'shed light on the structure of the L2 lexicon. The quality of interlanguage performance is generally assessed in terms of communicability, so that if an utterance communicates well, its quality will be assessed positively. Considering this, the more LEs a text displays, the less effective its communication and, therefore, the lower its quality' (Engber, 1995 in Llach, 2011:xii). LEs reduce the perceived quality of a piece of academic writing significantly (Astika 1993; Ellis, 1994 and Engber, 1995 in Hemchua and Schmitt, 2006).

LEA can also inform SLA research. Llach (2011) found that there were reductions in LEs as learners progressed over the years, that LEs had a clear detrimental effect on the quality of student writing, and they accounted for approximately 30% of the variance. She discovered that as learners became more proficient, they made more semantic than form errors. She also uncovered a low but significant correlation between written LEs and measures of receptive vocabulary knowledge. Further, research into LEs may shed light onto how words might be perceived (L2 studies) (Carter, 2012:185).

Previous research into LEA has produced important findings with some consistency. Word choice and word form errors were responsible for over 45% of errors made by second language users in a study by Foin and Lange (2007). Reid

(1997), Ferris, (2009), Frodesen, (2009), Frodesen and Starna, (1999) found that L2 learners made errors with idiomatic expressions, and incorrect register in lexical choices, word choice, word form, and prepositional phrase errors.

James' (1998) justification for a focus on LEs has prompted a focus on this area in recent years (Yang and Xu, 2001; Hemchua and Schmitt, 2006; Llach, 2011; Picot 2017). The problem, as Kalkvist (1998:82) states, is that different researchers investigating LEA have adopted different frameworks, or taxonomies for analysis, making comparison of studies very difficult. However, see Study 1 for a replication of Hemchua and Schmitt (2006).

2.5.6 Definition of a lexical error

According to Llach (2011), many studies start from a definition of what the authors interpret as an LE. However, different authors have different opinions of what constitutes an LE, and this has given rise to a great variety in taxonomies for analysis over the last 60 years. This thesis uses a definition based on Llach (2011), George (1972) and Lennon (1991), and defines an LE as a deviation in form and/ or meaning of a target language lexical word [or multi-word unit] which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's expert language user counterparts when communicating in that same variety of English.

Distinguishing between lexical and non-lexical errors

As already discussed above in Section 2.4.2, one issue with EA is identification of errors, and separation of lexical and grammatical errors can be problematic. One way to identify a written LE could be to identify all errors and then separate what constitutes a grammatical error. Apart from grammatical, punctuation, and organisational errors, all other remaining written errors will be lexical in nature (I include errors of coherence and style under the heading of LE). This would include tonal and pragmatic errors.

It is interesting to take into account existing theories of language when considering what constitutes grammar and lexis before distinctions between the two can be made for LEA purposes. The decisions for including and excluding particular types of errors in the LEA are based on the following scholarship on what constitutes lexis and grammar. My initial teacher education taught me that lexis was content words, such as nouns, verbs, adjectives and adverbs, and grammatical words

were pronouns, conjunctions, articles and determiners (Harmer 1991). Chomsky (1957) originally saw grammar and lexis as two separate systems with syntax independent of semantic meaning, and with his theory of Transformative Generative Grammar stated that language was made up of a finite set of grammar rules and 'native speakers' can create an infinite number of utterances, based on those rules, with lexis slotted in. Leech (1982: 179) described the 'closed' system of grammar as 'that part of a language that can be described in terms of generalisations or rules' and lexis as 'all the particular facts about language...those which cannot be generalised into rules', an 'open' system. Later a more pragmatically oriented theory of language, Systemic Functional Linguistics (SFL), was devised by Halliday (1985). He stated that SFL is a usage-based, social semiotic system and language was still a grammar, but incorporates lexis. From the perspective of SFL, language is all about choice in terms of the words one uses to represent the world. SFL gave rise to frameworks such as the lexicogrammatical approach (See Section 2.5.3), which suggested that a distinction between grammar and lexis is problematic, that the boundaries between the two systems are blurred and that they are integrated and interdependent. According to systemic functional theory, lexicogrammar is diversified into a metafunctional spectrum, extended from grammar to lexis, and ordered into a series of ranked units (Halliday 2013). James (1998: 143) states that 'Morphological aspects of words, which used to be treated as part of grammar, can just as well be treated as part of the word'.

The issue of collocation further blurs the boundary between lexis and grammar and compounds the problem of separating lexical errors from grammatical ones. However, collocation is commonly discussed in terms of lexis (e.g. Wray 2008), just as 'pattern grammar', a similar concept, has been viewed as belonging to the lexical syllabus by Hunston *et al.* (1997). To use collocations accurately, learners are required to have intuitions about lexical patterning and these stem from grammar, semantics and register and may even constitute a discrete type of knowledge (Halliday 1966). Hunston *et al.* (1997) use the term 'pattern grammar' when describing their approach to teaching lexis and grammar as interconnected language. Their research is based on working with large corpora and to exemplify this, they use examples of verbs that must be followed by the bare infinitive (e.g. appear and manage) and those that cannot (e.g. suggest and finish). Therefore, it was strongly felt that LEA should include a focus on language at the phrase level.

Despite these, at times, conflicting views of what constitutes grammar and lexis, if one were to focus on lexical only errors in an LEA, one would still require a systematic way of separating lexis from grammar. One method of doing so is provided by Hemchua and Schmitt's (2006) identification of lexical error guidance. They listed what constitutes a grammatical error and suggested all other errors could be considered lexical. This guidance was followed in NewLEAF 1 and 2 (see Appendix 6.7) and given some clarification in NewLEAF3 to make it more user-friendly (See Appendix 7.16). Some inclusions in the framework, such as formal elements, i.e. suffix, prefix and spelling, were included following Laufer's (1991) 'synformic confusions' (Hemchua and Schmitt 2006). The decision was taken to include all prepositions (also following Hemchua and Schmitt 2006) and all conjunctions, not just because including these in the analysis would make the exercise more useful through their inclusion, but also because, despite the fact that they have traditionally been seen as within the realm of grammar, they do carry semantic meaning and in my view are more towards the lexical end of the lexis/grammar continuum. For further justification of the inclusion of the different categories of lexis and therefore lexical errors in the NewLEAF frameworks, see Section 5.1.2. Reasons for conducting a lexical only error analysis can be found in Section 1.2.

There are other subjective potential issues in the identification of LEs:

Style – Readers of different genres of writing will have certain expectations of style in terms of lexicogrammatical choices. Few would disagree that 'kids' is a lexical error if it appears in an academic essay, for instance, but not all words can be unambiguously categorised as formal or informal. An academic essay would be the most suitable sample to use in testing any framework and guidance in terms of error identification as it is supposed to be formal in nature. However, reference to the first person, for example, may be considered by some as an LE in academic writing but not by others.

Compound noun issues - Compound nouns seem to go through a process of being two words, then one hyphenated word and then the hyphen disappears. (Tooth brush, tooth-brush, toothbrush). This may just be changing style of editing and book production, but deviations from a perceived norm can be considered a spelling issue, especially with compounds: variants and evolution include: car park, carpark/ bus stop, bus-stop / web site, website.

Coinage – Errors, nonce words, borrowings, coinage, blends, and calques are, amongst others, established ways in which new words come into English (Bodley, 2016). For example, 100 years ago a teacher may have corrected ‘pizza’ to ‘cheese pie’. This asks the question ‘Does the utterance contain an error or a neologism?’ Answers could be quite subjective. Grammar errors can be seen as a more finite group, whereas lexis is more dynamic, expanding more than the grammatical system.

Artistic License- Creative writers can create new terms. It may seem unfair to take this license away from our learners and classify their clear and creative uses of the language as errors.

2.5.7 Taxonomies in the published LEA frameworks

There have been various methods of classifying learners’ LEs or taxonomies, some more detailed than others. Some focus more on the cause of error than linguistic description or lexical competence, and some focus more on modification (or as Corder named it, surface error taxonomy: omission, over-inclusion, misselection, misorder and blend). This section will briefly describe the types of LE taxonomies, according to Llach (2011), before examining some different LEA frameworks that have been used since the 1960s. In Llach’s own estimation, hers is a problematic system of analysis as so many taxonomies have overlapping features. Following each type below, she lists the types of LEs that the framework seeks to collect.

- 1) **Form- and content-oriented.** These include confusion of two formally similar words, wrong word formation, word invention, borrowing from L1, super-imposition of features of the L1 onto the L2, linguistic calque, omission, addition, wrong ordering of letters within the word, wrong choice of two similar words, hyponyms instead of superonyms, collocational errors and confusion between two semantically similar words. This seems to be the most common method of classification and some studies using this method have proved to be the most exhaustive ones (Hemchua and Schmitt, 2006, James, 1998 and Zimmerman, 1987).
- 2) **Description only.** These include approximations, hybrids or blends, field errors, sense relation errors, stylistic errors, connotative errors. This type of evaluation avoids problems associated with cause or source of errors and

focuses solely on describing the errors and classifying them. They are more useful for multi-national groups of learners or when the analyst is not bilingual. It is my opinion that trying to allocate cause to errors is too subjective to be reliable and I support the use of description only frameworks. However, allocating causality after classification of errors using description only may be useful in some situations, i.e. when there is a monolingual group with a bilingual teacher, as this will help raise awareness of some errors.

- 3) **Etiological.** The study of the cause of the errors or the mental processes behind the error. Wrong choice of individual words, wrong choice of combinations, spelling errors, derivational errors, semantic infelicity, transfer, borrowing, wrong derivation, conceptual confusion.
- 4) **Origin of influence.** L1 derived (transfer, borrowing), L2 derived (overgeneralisation, confusion) or teacher or materials induced.
- 5) **Linguistic.** Stylistic errors, syntactic errors, order errors, semantic errors, system errors, idiosyncratic errors.
- 6) **Word-class.** Noun errors, verb errors, adjective errors, adverb errors.
- 7) **Product-oriented.** Focus is on the product, the LE.
- 8) **Process-oriented.** Focus is on the process that leads to the LE.

This classification of frameworks is of some use, as it demonstrates the breadth of LEA. However, as can be seen from the list above, there is, unsurprisingly, much repetition within these systems of classification and also in the naming of some of the categories. The main types appear to be those which seek to find either the type or cause of error, or both. Herein lies one of the issues with EA, separating cause and type of error, the former being quite speculative. If one is to evaluate the usability and depth of analysis produced by these frameworks, one must look in depth at some of the studies that represent these areas above and issues that are encountered when using the frameworks. This section will next report on the most prominent LE studies and their frameworks in chronological order (including those LE studies done within wider EA studies), that have been published within the last sixty years, namely, Dušková (1969), Corder (1973), Zimmerman (1986), Engber (1995), James (1998) and Hemchua and Schmitt (2006). These were selected as they provide a mixture of types, as outlined by Llach above. Interestingly, the categories in the different frameworks reveals the authors' different views of what LEs are and their varied opinions on identifying the cause

of LEs. Their frameworks will be described below, and they will be evaluated in terms of potential ease of use (how easy it would be to classify all errors and avoid issues of dual classification) and depth of analysis (the useful detail of the results or the spectrum of errors that the framework reveals). When explaining the frameworks, examples are taken from the publications.

2.5.7.1 Dušková (1969) Etiologic/process-oriented

This study involved a wider EA (including grammar) of the compositions of 50 Czech postgraduate students. It investigated the writing with a limited number of categories, and only four for LE (See Table 3), and attempted to describe causes.

Interestingly, Dušková did not include direct L1 transfer as an error. She found that in her wider general EA, 'a considerable number of errors could not be classified at all' (1969:15). This could be explained by the limited number of categories in her framework. She found that 5% of the total number of errors (which includes grammatical errors) were made with prepositions. When considering LEs alone, she found that Category 1 accounted for 21%, Category 2 accounted for 12%, Category 3 for 26% and Category 4 for 11% (see Table 3 below). She separated non-systemic, or 'one off/nonce errors' into a separate category, which accounted for 30% of all LEs. Unfortunately, thorough comparison of results with other LEA studies, e.g. Hemchua and Schmitt (2006) is not possible, given that she categorised many of what are currently understood to be LEs as grammatical ones (i.e., prepositions and modal verb choice).

- 1) Confusion of words due to formal similarity (i.e., words that look the same, e.g., than/then, think/thing, role/rule, take part/take place, suppose/suggest). The focus here is on L2 confusion.
- 2) Confusion of words due to relatedness of meaning (i.e., words that have similar meanings, but are not synonymous, e.g., institution/institute, latter/last, interesting/interested, usage/use, lie/lay, definite/definitive, in the last/in the least and at first/for the first time). The focus here is also on L2 confusion.
- 3) Assumed equivalence (i.e., an L1 word may have several equivalents in English, e.g., do/make, way/journey, repair/correct, include/involve, attend/visit, next/further/other). The focus here is on L1 transfer.
- 4) Distortions (i.e., words that do not exist in L2, but are clearly based on the correct word choice. E.g., **evolute/evolate* for evolve, **eluciete* for elucidate, **physist* for physicist, **realants* for relations). The focus here is on L2 confusion.

Table 3 Dušková's (1969:35) Framework for EA.

Category	Percentage of total LEs
1 Formal similarity	21
2 Relatedness of meaning	12
3 Assumed equivalence	26
4 Distortion	11
Other nonce mistakes	30
Total	100

Table 4 Distribution of LEs in Dušková (1969:35)

Potential ease of use

With only four categories, LEA should be quite straightforward using this taxonomy, as they seem quite distinct from each other at first glance. Dušková includes multi-word units in her list of examples, which will make categorisation easier. However, problems can be envisaged with Category 1: it may be difficult to decide whether an error is semantic or formal. One might need to ask the writer about the meaning behind the lexical choice, making the LEA process more complicated. Dušková's last example of suppose/suggest in the [Confusion of words due to formal similarity] category above stretches a view of formal similarity. There could be disagreement between analysts on where this similarity may lie, which may lead to unreliable categorisation. Category 2 could act as a catch-all for derivative errors and near-synonyms: straightforward to classify, but of limited use in terms of depth of analysis. Errors in Category 3 may not be spotted by teachers of multilingual groups, making this framework only of use to single L1 groups with a fluent bilingual analyst. Problems may occur with Category 4 when deciding whether the error is in fact based on an actual L2 word: the example given, **realants* seems to be quite a serious distortion that some analysts may not spot/understand as 'relations' or allocate to this category. The final category might be used by analysts as a miscellaneous or catch-all area for those LEs that do not seem to fit into the first three areas.

Potential depth of analysis

This is obviously a limited taxonomy in terms of number of categories and would probably be unsuitable for an in-depth analysis of the types of LEs that students make. As a framework for LEA, it lacks important categories, such as [Collocation] and [Connotation] error, and there is no provision for [Missing] or [Extra] lexis. On the whole, this framework may not offer sufficient detail in terms of the types and frequencies of LE, but would probably account for many LEs that a student would make. Presenting the results to learners and teachers may enable them to see

some of the more frequent and serious error types, but would be of limited use for the SLA researcher investigating lexical acquisition.

2.5.7.2 Richards (1971) Origin of influence criterion

Richards' (1971) study was designed to highlight the inadequacies of Contrastive Analysis, a popular method at the time of writing for predicting the type of errors made (see Section 2.3.1). By collecting language data from a wide range of L1 groups, he proved that the language errors he collected were not due to L1 interference, but were caused by 'the strategies employed by the learner in language acquisition, and from the mutual interference of items within the target language' (1971:22). Richards does not attempt to quantify the grammatical or LEs that he found. His presentation of the types of errors are revealing but the categorisation of the LEs seems problematic in that the taxonomy does not appear to be complete. Richards' framework sought to identify both grammatical and LEs, so it is not wholly applicable to LEA alone, but the framework does warrant some discussion. Although his categories allowed more for a grammatical focus on error identification, some may have been applicable to lexical production. Category 1 clearly relates to grammatical errors, although one could envisage situations where a learner might coin a new term based on her current lexical knowledge and desire to communicate, and transgress the rules of lexis. Category 2 looks problematic when allocating errors of this type. The analyst would have to speculate how the error was caused (what the original analogy might have been). Category 3 also looks problematic: the lexical example given was the only one in a series of grammatical errors exemplified. Identifying the cause of such an error would be problematic for the analyst, which would have to be done in order to allocate it to this category. This speculation could be too unscientific for accurate LEA. Allocation of error to Category 4 also seems problematic for the same reasons as those given above: identification of cause of error may just be speculation (I discuss below the usefulness of the categorisation of commonly confused lexemes).

- 1) Overgeneralisation (Covers instances of where the learner creates a deviant structure on the basis of his experience of other structures in the target language. E.g. *He can *sings.*)
- 2) Ignorance of rule restrictions (Failure to observe the restrictions of existing structures, that is, the application of rules to contexts where they do not apply. E.g. '*He said to me..*' could be used, through analogy to

generate the error '*He asked *to me....*' Or '*The rise in temperature...*' could cause the error '*The mercury *rise up the tube*'.)

- 3) Incomplete application of rules (The occurrence of structures whose deviancy represents the degree of development of the rules required to produce acceptable utterances. E.g. Question: *Do you read much?* Answer: *Yes, I read *much.*)
- 4) False concepts hypothesized (Derived from faulty comprehension of distinctions in the target language, which may be caused by 'poor gradation of teaching items' E.g. errors with *too*, *so* and *very*, *come* and *go*, etc).

Table 5 Richards' (1971:7-15) Framework for EA

Richards presented his results by categorising the errors he found into tables in the appendices of his paper. These sub-categories were as follows:

1) Errors in the production of verb groups, i.e.:

Be+verb stem for verb stem (e.g. *We are *live in this hut*)

Be+verb stem+ed for verb stem+ed (e.g. *Farmers *are goed to their houses.*)

Wrong form after do (e.g. *He did not *found.*)

Wrong form after modal verb (e.g. *Can be *regard as...*)

Be omitted before verb+stem+ed (particle) (e.g. *He *born in England*)

Be omitted after be+participle verb stem (e.g. *The sky is *cover with clouds.*)

Be omitted before verb+-ing (e.g. *They *running very fast*)

Verb stem for verb stem+s (e.g. *He *come from India.*)

2) Errors in the distribution of verb groups

be + verb + -ing for be + verb + -ed (E.g. *I am *interesting in that.*)

be + verb + -ing for verb stem (E.g. *She *is coming from Canada*)

be + not + verb + -ing for do + not + verb (E.g. *I *am not liking it*)

be + verb + -ing for verb +-ed in narrative (e.g. *In the afternoon, we *were going back.*)

verb stem for verb + -ed in narrative (E.g. *There were two animals *who do not like each other*)

have + verb + -ed for verb + -ed (E.g. *He *has arrived at noon.*)

have + be + verb for be +verb +-ed (E.g. *He *has been married long ago*)

verb (+-ed) for have + verb +-ed. (E.g. *We *correspond with them up to now*)

be + verb +ed for verb stem (E.g. *This money *is belonged to me*)

3) Miscellaneous errors

Wrong verb form in adverb of time (e.g. *I shall meet him before the train *will go*)

Object omitted or included unnecessarily (e.g. *we saw him play football and we admired **)

Errors in tense sequence (e.g. *He said that there *is a boy in the garden*)

Confusion of too, so, very (e.g. *I am *too tired that I cannot work*)

4) Errors in the use of prepositions

'With'

'In'

'At'

'For'

'On'

'Of'

'To'

All used instead of no preposition or another preposition.

5) Errors in the use of articles

6) Errors in the use of questions

Table 6 Richards' (1971:24-35) Framework for EA: Further sub-categorisation

The six categories above were never meant to be used as a framework for LEA alone, but 2, 3 and 4 do make for some interesting proposals for LEA.

Potential ease of use

Having categories for [Extra] and [Omitted word/words] will be helpful in categorising LEs. However, the metalanguage used above to detail the linguistic categories may cause more confusion for teachers who have little training in linguistic description of language. Other frameworks, with less detail, use less metalanguage.

Inclusion of the group [Confusion of too, so, very] (e.g. *I am *too tired that I cannot work*)' in 3 presents an interesting proposal for the categorisation of LEs.

Presentation of this type of error to students would be of value, and many teachers would report seeing this type of common error made. However, if one were to include categories of commonly confused, specific words, (personnel/personal, this/that, much/many), there would be too many categories to deal with LE in this way. Furthermore, this could cause dual allocation to category when considering personnel/personal as a spelling error. Also, inclusion of a [Miscellaneous errors] category seems too general to be of use to the lexical error analyst performing LEA for SLA studies purposes, but would serve as a useful place to ensure that all errors are categorised and not excluded.

As stated above, allocating some errors to some of the categories would require some speculation on the part of the analyst, revisiting the problematic area of dual categorisation of errors. Also, a degree of linguistic knowledge would also be required to confidently and consistently categorise the errors in this framework.

Potential depth of analysis

What is immediately clear is that Richards' original framework is a very limited taxonomy and would not provide a very detailed or useful analysis of lexical ability. According to Hemchua and Schmitt (2006) and the results of Chapter 3 in this

dissertation, errors in the use of prepositions are one of the largest groups made by L2 learners in terms of number of errors made. Interestingly, Richards omits a category for [Omission of prepositional partner]. However, having this breakdown of errors made with specific prepositional partners would make very useful corrective feedback for L2 learners. If the tables Richards presented in his appendices were to be used as a framework for LEA, they would have to be further developed. For example, Category 4 above would have to include all possible prepositions with which an error could be made, i.e, against, along, etc, etc so that all errors could be categorised.

Although the majority of error types in group two are more grammatical in nature, the first seem to be more lexical. This is supported by Lewis's (1992) definition of language as grammaticised lexis, as opposed to lexicalised grammar.

2.5.7.3 Corder (1973) Grammatical or linguistic criterion

What Corder's framework seems to lack in depth of analysis, it gains in ease of use. The first two columns ([Orthographic/phonological] and [Grammatical]) can be omitted for LEA study. The framework is relatively straightforward and self-explanatory:

	Phonological/orthographical	Grammatical	Lexical
Omission (an item has been missed out)			
Addition (an item has been included unnecessarily)			
Misselection (the wrong item has been chosen)			
Misordering (the items are in the wrong order)			

Table 7 Corder's (1973:278) Framework for EA

Potential ease of use

Because of this limited number of categories, difficulty of allocation of error is not predicted to be high. Considering Corder's original framework, although there is a clear link between phonology and spelling, having only one column for [Phonological/orthographical] errors may not allow for separation of other orthographical errors (e.g. [Punctuation], [Word breaks], etc). Surely it would provide more clarity and better organisation to separate the two. It is not clear whether Corder suggests using the table as a tally sheet for instances of errors for a group

of students or whether the table should be used to record an individual's errors and/or the reconstructions that he discusses. If the former, the framework table would not allow for a detailed picture of types of errors to be gained. If the latter, it would make for a better understanding. In 1973, Corder advocates the use of this framework and offers some guidance for its implementation. Unfortunately, he does not apply the framework to any language data, and is therefore unable to provide any findings. However, later in his chapter (1973:275), he states that once errors have been identified and corrected, they need a linguistic description. This extra stage would allow for more depth of analysis, but would draw out the process considerably and take a step away from quantitative recording of errors.

Potential depth of analysis

As can be seen from the framework above, Corder's (1973) framework does not go into much detail in terms of type of [Lexical] (or [Phonological/orthographical] or [Grammatical] error) categorisation and is limited to [Omission], [Addition], [(mis)Selection] and [Ordering] error in three/four language system areas. Having more categories for classification in the framework would make the analysis more detailed.

2.5.7.4 Zimmerman (1986) Descriptive/product-oriented

Zimmerman sought to create a more detailed classification of the LEs in the written work of German learners of English using a quite comprehensive taxonomy with 11 categories. The columns on the right are for tally marks when performing EA on a number of scripts.

Potential ease of use

Zimmermann states (1986:31-32) that the classification of error by explanation/cause is problematic and EA should focus only on the difference between the interlanguage lexeme and the intended target language lexeme. He disregards the learning process. An explanation of the cause of the error is useful to students, teachers and SLA researchers, as it accounts for, and may help to correct, many errors. However, I must also agree that the consideration of the reasons behind the error relies on some speculation and can result in dual classification of error. Therefore, an easier and better procedure would indeed be to identify error first, classify it based on its deviation from the target norm, and later seek to allocate causality, if possible, as suggested by Corder (1967). According to Llach (2011), errors could be allocated to more than one category, depending on the analyst. One could predict problems of dual categorisation

between 9 and 5, 6, 10 and 11, for example. Zimmermann himself admits to dual categorisation possibilities when allocating error to some of these categories. Indeed, to consider various sense relations for many errors is quite a time-consuming and problematic endeavour, especially for teachers who lack linguistic training.

Grouping	Sub-type, explanation and examples	N	V	Aj	Av
	<p>1) Sense Relations Errors (Errors made with confusions between superonyms, hyponyms, cohyponyms, heteronyms E.g. <i>I was thrilled by the special *smell*</i>. <Scent)</p>				
	<p>2) Field Errors (Errors made by using a word from the wrong field E.g. <i>...in a memorial* by the Foreign Office</i>. Instead of using 'memorandum', a word from the semantic field of written public communication, the learner wrote <i>memorial*</i>, which is from the field of public buildings.)</p>				
Componential meaning error	<p>3) Feature Errors (Errors that are described through partial aspects of meaning or meaning components or semantic features. E.g., <i>I slided* and fell</i>. <Slipped)</p>				
	<p>4) Word-Formation Errors (Errors relating to the spelling or morphology of lexemes: formal and semantic. E.g., <i>The paper is *yellowy</i>)</p>				
General/Other	<p>5) Collocation Errors (Errors that involve words that co-occur frequently)</p>				
	<p>6) Idiomatic expression Errors. (Errors with omission of or confusion with idioms E.g. <i>These negotiations are cutting a great figure*</i> >playing an important role.)</p>				
	<p>7) Omission Errors (Errors that involve a missing word/words e.g., <i>I had bitten through *_____ my tongue</i>. >the edge of)</p>				
	<p>8) Redundancy Errors (Errors that include an unnecessary word. E.g., <i>At a rate* speed of five miles per second</i>.)</p>				
	<p>9) Paraphrase Errors (Errors where more than one word requires replacing with possibly more than one word. These errors could be formally possible or formally deviant E.g. <i>The *outer parts of my tongue</i>. >Edges. <i>And the tongue was a muscle which could not be missed*</i>. >The Tongue is muscle that is indispensable.)</p>				
	<p>10) Stylistic Errors (Errors using an incorrect level of formality or register. E.g., <i>My dealing with written *stuff</i>. >Things)</p>				
	<p>11) Connotative errors (Errors that invoke the incorrect idea or feeling. E.g., <i>I *jerked upstairs</i>. >Jumped.)</p>				
Uncat					

Table 8 Zimmerman's (1986:32-36) Framework for LEA

Having a category titled [Word-Formation Errors] (errors relating to the spelling or morphology of lexemes: formal and semantic. E.g. *The paper is yellowy**) section

solves potential problems when trying to decide whether an error is a spelling error or a made-up word or a word family error.

Potential depth of analysis

Zimmerman (1986) sought to differentiate LE by word class: [Noun], [Verb], [Adjective], [Adverb] and type of errors made with these types of words against the type of error that can be made with them, i.e. [Sense Relations], [Field], [Word-Formation], [Collocation], [Redundancy], [Omission], [Paraphrase] and [Connotation]. This would provide a clearer picture for LEA than simply counting the number of errors made using either axis (See Table 8 above.).

The 11 sections appear to cover many areas that previous frameworks have omitted and the dual axes allow for much depth of analysis (e.g. The analyst could now identify whether a [Stylistic] error had been made with an incorrect [Noun], [Verb], [Adjective] or [Adverb], supplying a much more useful picture for the learner, teacher and SLA researcher). The inclusion of a [Paraphrase] error category is an effective idea that deals with issues in the categorisation of errors where a whole clause or phrase is deviant, where no single lexeme could be identified as problematic on its own and the whole clause could seem erroneous. The suggestion of rewriting the clause overcomes this problem. Lennon (1991b) also makes a similar suggestion in terms of identifying the domain and extent of an error (see Section 2.4.2). Further, having a [Paraphrase] category allows for avoidance of correcting individual, multiple lexemes in a phrase when the corrected version is still 'unnatural'.

Zimmerman, aware of the issues of categorisation, tentatively reports on tendencies, as opposed to results. Analysing his own corpus of learner errors, he reports that 20% are [Sense relations] errors, approximately 33% are [Field] errors, 10% could be seen as [Feature] errors and 20% as [Wrong collocations]. Overall, the framework seems promising in terms of ease of use and depth of analysis.

2.5.7.5 Zimmerman (1987) Form-and content-oriented/product-oriented

Zimmerman analysed translations from German to English by advanced German learners of English to ascertain the type of LEs they made. He describes not a framework/table that can be implemented, but a concept of analysing LE on the

basis of form and content. He does not quantify his findings. The article focuses on differentiating between the two. He categorises the errors as follows:

Form-orientation

- 1) Phonic/graphic L1 orientation
- 2) Phonic-graphic L2 orientation
- 3) Indirect form-orientation
- 4) Phonic-graphic orientation ambiguous between L1 and L2
- 5) Morphological L1 orientation
- 6) Morphological L2 orientation
- 7) Phrase level form-orientation

Content-orientation

- 8) Sense-relations
- 9) Semantic frames

Table 9 Zimmermann's (1987:58-62) Framework for LEA

Potential ease of use

According to Llach (2011) the main problem with Zimmerman's work is 'the frequent lack of clear definitions, the abundance of vague explanations, and the mixing of classification criteria (psycholinguistic cause of the error, linguistic source of the error (L1 or L2) and product orientation. However, the framework, mainly due to its originality of analysing form and content errors, is still of interest. Zimmerman states that differentiating between [Form oriented] and [Content oriented] LE is problematic, e.g., *schrammte* =scraped or brushed alongside. Learners produced *scrammed*, *scrabbed*, *scratched*, *touched*, *got at* and *bumped into*. This range of errors clearly go from [Form] to [Content orientation], but categorising some in the middle is problematic. He demonstrates this with a variety of examples of German learners' of English errors. He states that '[Form-] and [Content-orientation] occur at different levels of language, that it (the error) can be L1 or L2 oriented, that it is a matter of degree, and that there can be instances of mixed sources' (1987:55). He continues to offer guidance for the differentiation between [Form-] and [Content-oriented] error. However, he also states that differentiation is, at times, still problematic. Furthermore, all but an expert linguist who is proficiently bilingual would struggle to understand the differences between the categories. Obviously, the analyst would have to be proficiently bilingual in order to confidently identify interlingual errors, making this framework less useful where teachers do not speak the learners' L1.

Potential depth of analysis

The number of categories allows for an in-depth analysis, but despite this,

confident allocation of error to category would still be problematic, as reported by Zimmerman himself throughout the article. Despite detailed categorisation in the areas that Zimmerman proposes, the framework lacks categories such as [Cohesion], [Coherence], [Collocation], etc. It mainly focuses on the differences between L1 form oriented, L2 form oriented, L1 and L2 content-oriented errors.

2.5.7.6 Meara and English's Framework (1987)

<p>0) Totally wrong word E.g., <i>You would *supply to change it</i></p> <p>1) Phonologically related word E.g., <i>no sign of a *punch on the tyre</i></p> <p>2) Wrong word, right semantic area E.g., <i>my wife *tranquilised me</i></p> <p>3) Formal derivational errors E.g., <i>there is an *amusing arcade</i></p> <p>4) Usage E.g., <i>he easily accessed drugs thanks to his money</i></p> <p>5) Spelling error [where this resulted in another word] E.g., <i>the *prize of the book was two pounds</i></p>

Table 10 Meara and English's (1987:4) Framework for LEA

Meara and English sought to examine the effectiveness of English dictionaries in assisting error correction by group of mixed L1 beginners. They found that analysis of errors made by learners can be useful to dictionary writers to improve the dictionaries and that different dictionaries vary in effectiveness for different groups of learners. It focusses mainly on form and semantic errors. The most common error was [Wrong word, right semantic area], followed by [Usage] and then [Totally wrong word]. Percentages of total errors made were not mentioned.

Potential ease of use

These are clearly named categories and allocation to them should not be problematic. The [Usage] category could act as a catch-all for [Incoherence] and [Infelicitous phrase] errors. Meara and English also avoid problems of identifying cause. Inclusion of Category 0, [Totally wrong word] is sensible, as this category is lacking in many other frameworks, although one could predict problems categorising a word which could be considered as such or simply spelled incorrectly.

Potential depth of analysis

Given the lack of categories, using this framework would not provide a broad spectrum of errors. It is notable for a lack of focus on fixed expressions, and the surface error taxonomy first introduced by Corder (1967).

2.5.7.7 Lennon (1991) Grammatical or linguistic criterion/word-class criterion/product-oriented

Although designed to analyse a spoken error corpus compiled from five of his advanced German learners, Lennon's framework could also be adapted to analyse written work. Lennon reports it to be an exhaustive taxonomy with mutually exclusive categories and claims that only 1-2% of errors could not be classified. Lennon found that the LEs made up 23% of all errors, but does not offer a further breakdown of sub-type. Again, Lennon chooses to avoid identifying causes.

- 1) Intra-lexeme
 - a. Phonological error
 - b. Verb morphology error
 - c. Noun morphology error
 - d. Adjective and adverb morphology error
 - e. Categorisation error
- 2) Intra-NP
 - a. Initiator, determiner, adjectival choice error
 - b. Adjective + noun and noun + noun combinations
 - c. Noun phrase post modification error
- 3) Intra-verb group
 - a. Tense and aspect choice error
 - b. Choice of 'co-verbs' (modals and catenatives) and auxiliary and participle combinations (excluding tense choice)
- 4) Prep./adv. particle choice
- 5) Preforms choice
- 6) Adv./part. Position
- 7) Verb complement
- 8) Clause linkage
 - a. Conjunction choice
 - b. Relative pronoun choice
 - c. Omission of the second auxiliary in subordinate clauses
- 9) Sentence structure
- 10) Lexis
 - a. Verb choice errors without collocational inappropriacy
 - b. Verb choice errors with collocational inappropriacy
 - c. Noun choice errors without collocational inappropriacy
 - d. Noun choice errors with collocational inappropriacy
 - e. Adjective choice errors without collocational inappropriacy
 - f. Adjective choice errors with collocational inappropriacy
 - g. Adverb choice errors without collocational inappropriacy
 - h. Adverb choice errors with collocational inappropriacy
 - i. Miscellaneous collocational errors

Table 11 Lennon's (1991:34-39) Framework for EA

The framework's focus is on grammatical EA and part of speech, as opposed to categorisation of types of LE. Therefore, categories such as, [Coherence], [Verbosity] and [Overspecification] categories are missing.

Potential ease of use

It could be considered easy to use in terms of avoiding dual classification issues because it is easier to identify word-class and therefore allocate error to categories based around word-class alone. However, confident allocation to category may not be easy for the non-linguistic expert, due to the required knowledge of linguistic meta-language. Examples would help with this to some degree. Also, Lennon's framework makes no attempt to analyse the cause of error (intra or interlingual for example), but only sought to describe the errors made. This makes the process more straightforward. Also, the guidance he offers for categorisation helps the analyst achieve some consistency. The penultimate category, [Sentence structure], seems to be quite broad, which will also assist in the reduction of uncategorisable errors. However, with a blurred distinction between grammar and lexis, there could be occasions when the analyst could deliberate whether to allocate an error to Category 10, [Lexis], or Category 7, [Verb complement], for example.

Potential depth of analysis

The sheer number of categories allows for very good depth of analysis, providing a very comprehensive picture of the types of errors a learner may make. However, the framework does not refer to whether an error is made in [Omission], [Addition], [Selection], [Ordering] or [Blend], as James (1998) suggests. This would make the resulting analysis less clear for the learner or teacher seeking to improve language accuracy following the analysis. An extra step would be required to do this, making the process longer and more time-consuming. Section 10, which focusses on [Lexis], may present some issues in that some sub-categories are absent from the framework, including [Coherence], [Verbosity] and [Overspecification] errors, but Lennon did focus on advanced learners' work. Feedback in these types of areas, as opposed to part of speech, provides better feedback for the learner.

2.5.7.8 Zughoul (1991) Etiologic/product/process-oriented

Zughoul, in an attempt to provide a more detailed categorisation of the type and frequency of errors, often interference errors, that his Jordanian students made, devised a framework with 13 categories. See Table 12 below.

This is quite an original framework that caters specifically for Arabic learners. Highlighting many of these error types would be very helpful to students, especially Categories 1 and 2. Zughoul argues for the importance of Category 3, [Derivativeness], on the basis of the derivative nature of Arabic and how this would

lead to another type of interference error. He found that the most common LE was [Assumed synonymy] accounting for 24% of all LEs, followed by [Literal translation] (12%) and then [Derivativeness] (10%).

- 1) Assumed synonymy e.g. *There are many *works in the city* <jobs. This category is useful for identifying errors caused by the students' L1 containing a word which can be translated into two words in L2, but only one of these translations can be used in certain contexts (e.g., *tall man* and *long journey*). Bilingual dictionaries are often the cause of such errors.
- 2) Literal Translation e.g., in Arabic, 'compelling circumstances' translates into English as **Cairo envelopes*.
- 3) Derivativeness e.g., *Mansaf is the national *cook in Jordan* <dish.
- 4) Collocation e.g., *The weather is *kind in the country*.
- 5) Similar forms e.g., *People are unable to work and earn *efficient money.* >sufficient. These could be graphic or phonetic similarities. Both words exist in the L2.
- 6) Message translations e.g., *My room *reflects the relax on myself*. Here the learner provides a word-for-word translation of the L1 for the whole sentence. This differs from literal translation which is simply a phrase within a sentence.
- 7) Idiomaticity e.g., *I *sleep at 11 o' clock*.
- 8) Influence of Arabic style e.g., *My house is *placed or situated in a beautiful area*.
- 9) Paraphrase or Circumlocutions e.g., *There is a *clothes cupboard in my room*. This category is for multiple word verbosity.
- 10) Verbosity e.g., *Everyone of the participants *vanished*. (left) This category is for single word verbosity.
- 11) Analogy e.g., *At half past seven, I *supper*. These could be single or multiple-word items where the learner 'coins new verbs, nouns, adjectives, etc, along the lines of existing paradigms.
- 12) Binary terms e.g., *I returned the books I *lent from the library*.
- 13) Overuse of some lexical items e.g., overuse of words such as good, bad, big and small.

Table 12 Zughoul's (1991:47) Framework for LEA

Potential ease of use

Some of these categories seem quite vague and one immediately sees scope for overlap. Zughoul himself (1991:47) states that some errors can be categorised under 1-5 areas at the same time, making this one of the least reliable frameworks under the current analysis. This would make comparative studies rather unreliable.

Particularly problematic would be the distinction between 2, 6, 8 and 13. Again, the analyst would have to be a proficient bilingual speaker to use this framework. Category 4, [Collocation], is an important category to have, but the absence of categories, such as [Awkward expression] may lead to a lack of depth of analysis. Furthermore, it is quite subjective whether '*life is sweet in the city*' (another one of Zughoul's examples of errors in this category) is erroneous. It would also be possible to interpret some of these examples as [Assumed synonymy] or [Literal translation]. Category 7, [Idiomaticity], seems to be a strange title for this category, given the examples that Zughoul provides. They seem quite arbitrarily allocated. Category 8, [Influence of Arabic], seems problematic. Examples of errors in this category could equally be placed in several others. Category 9, unfortunately, also provides much potential for confusion between this category and several others. Allocation of error to 11, [Analogy], may also be problematic as these errors could also be entered into Category 1. One could also allocate errors to Category 3, for example. Category 12, [Binary terms], is more of a style issue than an error issue. Again, allocation to this category would be subjective, unless one considers all usage of these words as problematic.

Potential depth of analysis

There is a good number of categories, but to provide a more comprehensive picture of errors in a learner's lexicon, one would need to add [Coherence], [Cohesion] and [Spelling] as a minimum. Inclusion of [Interlingual errors] will provide some reasons for the cause of some errors, but having these as a category on their own will not allow for the description of the type of errors that are due to L1 interference. However, there is some originality in the framework.

2.5.7.9 Engber (1995) Etiologic/product-oriented

Engber (1995) sought to compare measures of lexical richness with teachers' ratings of composition quality. She found that totalling the number of error-free content word lemmas gave the strongest correlation (.57) with the teachers' ratings. She does not state in her paper which types of LEs were the most frequent. Despite Llach (2011) classifying this as an Etiological framework, it seems to follow the more common form/semantic framework.

Lexical choice

A) Individual lexical items

<p>1) Incorrect word choice– semantically unrelated (e.g., <i>It has some *meanings to study in the US.</i>)</p> <p>2) Incorrect word choice – semantically close (e.g., <i>We can study some *strange subjects.</i>)</p> <p>B) Combinations</p> <p>1) Two lexical items (e.g., <i>Young people can say their ideas</i>)</p> <p>2) Phrases (e.g., <i>We have a lot common: >a lot in common</i>)</p> <p>3) Multiple errors involving core lexical items (e.g., <i>It is being popular year and year</i>)</p> <p>Lexical form</p> <p>1) Derivational errors (e.g., <i>It keeps the class more activity</i>)</p> <p>2) Verb forms (e.g., <i>It isn't a good way for looking a job</i>)</p> <p>3) Phonetically similar, semantically unrelated (e.g., <i>I thing that English will help my country</i>)</p> <p>4) Word distorted-major spelling error (e.g., <i>We can see the roadshows stimulously – simultaneously</i>)</p>

Table 13 Engber's (1995: 146) Framework for LEA

Potential ease of use

Engber uses nine categories to analyse LE types and makes no attempt to describe the cause behind the errors in her taxonomy, which makes classification more straightforward. This decision not to focus on cause may be partly due to the fact that her participants were 66 students of mixed L1. Part 1B seems to be problematic at first glance. It seems odd to differentiate between errors with two lexical items and errors with phrases. Also, [Multiple errors involving core lexical items] may be confused with [Phrase] errors. More guidance is required here. B3 could be problematic in that it could be a matter of opinion as to how phonetically similar or semantically unrelated a word may be. Similarly, it could also be subjective as to how major a spelling error is.

Potential depth of analysis

Interestingly and surprisingly, Engber's is one of the few frameworks to include a 'completely wrong word' category. This is useful when distinguishing between completely wrong word and a semantically close word (Category 2). Also, the focus on chunks of language is in line with a modern view of lexis (Lewis 1992). However, the rather limited number of categories would not provide a satisfactory detailed picture of the spectrum of errors made.

2.5.7.10 James (1998) Form and content-oriented

Taken from his seminal book on EA, where he seeks to revive the practice, James'

1998 framework is not especially designed exclusively for LEA. However, his framework could be easily adapted for LEA by omitting some columns.

Level	Substance	Text		Discourse
	Phonology Graphology Spelling Pronunciation	Grammar	Lexis	Cohesion Coherence Genre-fidelity Felicity
Modification		RANK: Clause- Phrase- Word- Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc.	Sense relations Collocations	
Omission				
Over-inclusion				
Misselection				
Misorder				
Blend				

Table 14 James' (1998:274) Framework for EA

James advocates a dual system of classification: that of modification ([Omission], [Over-inclusion], [Misselection], [Misorder] and [Blend]) and level or language system, which focuses on: [Substance] ([Phonology], [Graphology], [Spelling] and [Pronunciation]): [Text] (including [Grammar] and [Lexis]) and [Discourse] ([Cohesion], [Coherence], [Genre-fidelity] and [Felicity]). It is quite an original and comprehensive framework, as James has included the category of [Blend] (two L2 words are combined to produce a new, erroneous word (e.g., **Travell*, from 'travel' and 'travelled')). He has also introduced a [Discourse] main category with ([Cohesion], [Coherence], [Genre-fidelity] and [Felicity] under this heading. Again, the blank cells can be used for tally marks if analysing a large number of scripts or the actual errors could be written into them if analysing just one sample.

Potential ease of use

Using a surface taxonomy with the addition of [Blend] appears to be a sensible method of categorising errors as the vast majority will fit into these categories and the possibility of dual categorisation is avoided here. This method also avoids

potentially difficult linguistic metalanguage with which not all potential analysts might be familiar. It could be that analysts will differ in terms of whether they allocate an error to [Coherence], [Genre-fidelity] and [Felicity], so grouping them together like this deals with this issue whilst still providing a category for these errors.

Potential depth of analysis

The dual axes and inclusion of [Blend] into a surface taxonomy in the horizontal axis is an intelligent addition, as otherwise, identification and categorisation of these errors would be quite arbitrary. Certainly, errors of [Coherence] should be viewed as the more serious error type. Including this category here will make the whole process of more value to the teacher and learner. Inclusion of the category [Felicity] will allow the analyst to report to the more advanced learner where their language is not so much wrong, but perhaps not what a more expert language user may have written. Identification of [Genre-fidelity] errors will allow for more refined lexical choices when working in different genres and would cater for verbosity. However, whilst the framework on first inspection looks as though it would cause few dual categorisation issues, there is not a wide range of LE type categories (e.g., [Verbosity], type of [Sense relations] errors, types of [Collocational] errors or types of [Discourse] errors). Interestingly, James maintains a more traditional view of the distinction between grammar and lexis with [Phrase] errors and [Conjunctions] on the grammar side of the fence.

In a case study to exemplify the use of the framework, he analysed a Brazilian's writing and found that the most common error type was [Misselection], suggesting that the learner has much language at his disposal, and that accuracy of language choice was his biggest issue. James concludes that noticing and focussing activities would benefit this learner most (James 1998:273).

2.5.7.11 Hemchua and Schmitt (2006) Form and content oriented

The most comprehensive taxonomy of LE types was produced by Hemchua and Schmitt (2006), based on Leech (1981), Laufer (1992) and James (1998) (See Table 15 below). Hemchua and Schmitt (2006) produced a framework to analyse the type, frequency and cause of LEs made by their university Thai students. This is clearly one of the more well-informed and comprehensive LE taxonomies, with 24 categories. Therefore, it should provide a very detailed picture of LEs made. However, some categories may not yield many errors, such as [B2.3 Arbitrary

combinations and irreversible binomials] or [B3 Connotation errors]. The framework divides errors into formal and semantic errors, which is a useful distinction. Hemchua and Schmitt found that the most common error was [B2.1 Semantic word selection] (31% of all LEs), followed by [B2.4 Preposition partners] (18%) and then [A2.3 Calque (translation)] (12%). Learners had more difficulties with semantics than the forms of words. Errors were mainly due to L2 difficulty, rather than L1 transfer issues. Interestingly, the authors did not draw from Lennon's (1991) and Taylor's (1986) recommendations for EA. Perhaps this was because they only focussed on lexis. Within the framework, Hemchua and Schmitt cited the authors from whom they gained their ideas for their various categories. The citations are included below.

A Formal Errors

1 **Formal misselection** (James 1998 drew from Laufer's 1991'synformic confusions' (Visual and sound similarity and named them malapropisms)

1.1 Suffix type (Correct root, but incorrect suffix, e.g., *They were very *considerable*)

1.2 Prefix type (Correct root, but incorrect prefix, e.g., *I'd like to *preserve a ticket*)

1.3 Vowel-based type (*Take a *set*)

1.4 Consonant-based type (*I need to *safe some money*)

1.5 False friends (*I feel *serious<stressed*)

2 **Misformations** (words that do not exist in English) Error source is L1.

2.1 Borrowings (L1 words) (*He was shot in the *kopf*)

2.2 Coinage (inventing based on L1) (*Smoking is *nocive to health*) (nocivo is Portuguese for harmful)

2.3 Calque (translation from L1) (*Find a car to bring us to *go to the hospital*)

3 **Distortions** (words that do not exist in English) Error source is confusion, not L1.

3.1 Omissions (**intresting*)

3.2 Overinclusion (**Dinning room*)

3.3 Misselection (**Delitouse< delicious*)

3.4 Misordering (**Littel < little*)

3.5 Blending (**travell <travel + travelled*)

Note that L1 inference (interlingual errors and interference within L2 (intralingual errors) can be at play at the same time in the same error. That is, L1 effects may lead to L2 effects.

B Semantic Errors (James 1998 proposes the first two types and sub-types)

1 **Confusion of sense relations**

1.1 General term for specific one (using a superonym for a hyponym, making meaning underspecified. *We have modern *equipment in our house.* <appliances.)

<p>1.2 Overly specific term (Using a hyponym for a superonym. <i>The *colonels <officers live in the castle.</i>)</p> <p>1.3 Inappropriate co-hyponyms (Inappropriate co-hyponyms, <i>The city has good *communication <transport, such as a lot of buses.</i>)</p> <p>1.4 Near synonyms (Using a wrong near synonym. <i>A *regretful <remorseful sinner.</i>)</p> <p>2 Collocation errors</p> <p>2.1 Semantic word selection (Semantically determined error. <i>The city is *grown <developed.</i>)</p> <p>2.2 Statistically weighted preferences (<i>The Army suffered *big <heavy losses</i>)</p> <p>2.3 Arbitrary combinations and irreversible binomials (<i>*hike-hitch</i>).</p> <p>2.4 Preposition partners (<i>surrounded *with <by nature</i>)</p> <p>3 Connotation errors (<i>He's quite *notorious for the charity work he's done</i>) (from Leech 1981)</p> <p>4 Stylistic errors (James 1998 from Leech 1981)</p> <p>4.1 Verbosity (<i>I *informed my girlfriend via the medium of the telephone.</i>)</p> <p>4.2 Underspecification (<i>Although *cars in the country are lower,....<Although there are lower car numbers in the country...</i>)</p>

Table 15 Hemchua and Schmitt's (2006: 12) Framework for LEA

Potential ease of use

In order to use Hemchua and Schmitt's framework, one would need to be bilingual or find assistance from bilingual researchers/analysts in order to confidently allocate to categories in Section 2. This section also re-introduces causality, which may be important when working with single L1 groups, but highly problematic with mixed L1 groups. Furthermore, it is highly probable that inclusion of causality would be behind many issues with dual categorisation, as errors can have both a cause and a description. The surface taxonomy is introduced in Section 3, which again will cause some dual categorisation issues. James' (1998) suggestion to have this in a different axis was not followed, but his suggestion to include [Blend] was. [B1.1] and [B2.1] could also cause some dual categorisation. Apart from these problems, other categories seem to be mutually exclusive and are clearly explained, with examples, making allocation to these categories easier for the less analytic linguistic expert. The authors offer some guidance in how to avoid inclusion of grammatical errors, which will help with error identification, but to overcome some of the possibilities for dual categorisation, the framework should be accompanied with some guidance that states which area to allocate certain errors to if they can be categorised in more than one.

Potential depth of analysis

This framework, which to some extent, builds on previous frameworks by including a wider variety of categories will provide great depth of analysis. Results will be suitably detailed to inform learners, teachers and SLA researchers. For these reasons, this framework was selected for a replication study (See Chapter 3 for a fuller account of actual ease of use and depth of analysis).

2.5.7.12 Llach (2011) Form/content/origin of influence criterion

One of Llach's (2011) primary interests was to identify causality of LE, as well as type. This is due to the fact that she was working with a monolingual group of young Spanish students of English. She also sought to ascertain whether there was an effect of LEs on assessments of writing quality, whether LE type and frequency changed over time and whether there were any correlations between LEs in compositions and knowledge of receptive vocabulary.

- 1) **Misspellings.** This category simply includes any type of spelling error.
- 2) **Borrowings.** This category is for 'complete language shift' or code-switching', where an L1 word or phrase is inserted into the L2.
- 3) **Coinage.** This category is for L1 words that have been modified in some way so that they look more like they belong in the L2. They are neither L1 nor L2.
- 4) **Calque.** This category is the literal translation of a word or phrase from the L1 to the L2
- 5) **Misselection** Based on Laufer's (1990, 1991 and 1992) malapropism or 'synforms', this category is for any confusion of formally similar items. The wrong word and target words exist in the L2.
- 6) **Semantic confusion.** This category is for confusion of semantically similar words. Again, two existing L2 words are confused, but they are similar in meaning, not form.

Table 16 Llach's (2011:123-124) Framework for LEA

She found: reductions in LEs as learners progressed over the years, that LEs had a strong effect on how writing quality is assessed and that learners made more semantic than form-based errors as they developed. She also uncovered a low significant correlation between written LEs and receptive vocabulary knowledge. She also ascertained that 'at grade 4, almost two-thirds of the errors were spelling mistakes and at Grade 6, just over half were misspellings' (p267).

Potential ease of use

The first category is easy to allocate errors to. No attempt is made in Llach's framework to attribute these to either L1 interference, inherent difficulty with the L2 (sound/spelling correspondences) or teacher/materials-induced error influence. Certainly, a proficient bilingual speaker will be able to identify and categorise errors made under headings 2-4. Differentiating between these would also be easy enough. It is possible that there could be the odd occasion that errors could be categorised as 5 or 6. Although not mentioned, incorrect affixation could be allocated to 5. Usefully, any error made with verb endings (or any word class) could also be allocated to Category 5, making this potentially usable for whole language, and not just LE classification. Collocation and connotation errors and preposition errors could be allocated to 6. Interestingly, Llach used a Pearson product-moment correlation coefficient test to check for interrater reliability when allocating errors to categories. A high score of 0.87 was found when 100 randomly selected compositions were analysed by another trained EFL teacher. However, perhaps this similarity is unsurprising, given the limited number of categories in the framework.

Potential depth of analysis

Three of the six categories relate to Hemchua and Schmitt's (2006) [L1 interference] categories: [Borrowings], [Coinage] and [Calque]. The remaining three, [Misspellings], [Misselection] and [Semantic confusion] will offer little scope for the identification of the full spectrum of LEs that have been seen in other frameworks. It is probable that the vast majority of LEs would fall into the last two categories: they would act as a catch-all and not offer much in terms of depth of analysis. The framework suffers from a lack of focus at a phrasal level. Nor does it separate out other useful categories, such as errors of [Cohesion], [Coherence], [Verbosity], [Underspecification], [Completely wrong word], etc. The ease of use that was achieved with relatively few categories would not allow for a wide distribution of error types and would therefore not provide the same breakdown of error types that the more comprehensive taxonomies might show (several studies have shown that errors with prepositions were amongst the most numerous). Perhaps Llach did not envisage many error sub-types in these categories due to the younger age and lower proficiency level of her participants.

This section has presented a description of the main frameworks for LEA published in the last sixty years, and has sought to evaluate them in terms of

potential ease of use and potential depth of analysis. The next section, the conclusion, will summarise the main points of the literature review with a focus on the main issues to be overcome if one is to produce a relatively easy to use framework for LEA that provides a suitably detailed set of results that will inform learners who wish to improve their English, teachers who wish to understand the types and frequency of LEs that their learners make, materials writers to produce materials that will help learners more and SLA researchers to better understand LEs.

2.6 Summary/conclusion

This literature review has examined what has been published in a range of subtopics related to LEA.

Section 1 provided a definition of errors in language teaching and learning and delivered the line of argument that there is still a need for LE correction in language teaching. This is largely driven by the focus on lexical accuracy in high-stakes tests, such as IELTS, perceived learner expectations for correction and to provide balance for the much-needed accuracy focused work promoted by proponents of more fluency-based approaches to teaching, such as task-based learning.

Section 2 introduced and defined EA in general and described a need for the procedure and its uses outside language teaching. It also outlined current limitations with a computer-aided EA: i.e. computers are still unable to identify and categorise errors to the same level of detail that a trained linguist would. Corpora still require human error 'tagging' before they can be used for analysis.

Section 3 contextualised EA by describing Contrastive Analysis, a system of linguistic analysis that came before it. Contrastive Analysis compared two languages to find their similarities and differences. Results were used to predict the areas in which students would make errors. CA proved to be inadequate, as it was shown that many errors that students made were developmental, and not just transfer errors. The section continued to offer reasons for the popularity of EA and offered reasons for its demise: procedural issues with some of the stages and a criticism of EA as a theory of language acquisition. This criticism is largely because the prevailing methods for language teaching at that time were grammar translation and audio-lingualism. As these methodologies became discredited, so did the associated EA in the quest for a more suitable teaching methodology.

Certainly some of the issues with the stages could be overcome, or at least mitigated, and a criticism of EA as a theory of SLA is well-founded, but does not warrant abandonment of EA as a tool to simply analyse student errors.

Section 4 examined the stages of EA (collection, identification, description, explanation and evaluation) and looked in further detail at issues associated with these stages. The two main stages of identification and description were of the most importance to successful EA and were also the most problematic. A rebuttal against the limitations raised about EA was presented and I concluded that EA required some modifications, but it remained an important and highly useful tool for investigating learners' interlanguage development.

Section 5 focused on LEA by defining it and stating the importance and centrality of lexis in language teaching and learning. It discussed a new view of lexis and grammar, lexicogrammar, and concluded that the two are difficult to separate for analytical purposes. A justification for LEA was provided before looking at the historic approaches to LEA and the potential that the different frameworks may have in terms of ease of use and depth of analysis. Given the evaluation in the previous section, it is clear that there is no perfect taxonomy of categories for LEA. It seems that there is a balance to be struck between ease of use and depth of analysis: too many categories could lead to dual classification issues and too few leads to restricted depth of analysis.

To summarise, it is my belief that LEA warrants further investigation, despite some issues surrounding the blurred boundaries between lexis and grammar, identification of errors and the categorisation of errors. I believe that it is a worthwhile endeavour to try to ascertain whether an improved framework can be found that overcomes some of the issues described above, even if there is still some overlap between the types of LEs that can be made, as an improved framework will cast more light on this important area and provide more comprehensive understanding of type and frequency of errors made. The first step in ascertaining the above would be to perform a replication study, using what looks like the most promising of LEA frameworks, Hemchua and Schmitt's (2006) framework. This is the focus of the next chapter and Study 1.

Chapter 3 (Study 1) Replication

3.1 Introduction

This chapter provides a replication of an LEA study and addresses a gap in the literature in lexical studies: more knowledge is sought to improve understanding of LEA procedures, and there are very few replications of LEAs that also seek to

establish what issues there may be with an existing framework. The few existing LEA studies were discussed in the previous chapter. After some decades when error analysis had fallen out of favour, Hemchua and Schmitt (2006) sought to create a new framework for LEA with the aim of providing a reliable and accurate procedure for establishing the type and frequency of LEs made in English language learners' writing. Using this framework, which added considerable value to earlier models (e.g. James, 1998), Hemchua and Schmitt categorised and counted the LEs in the written compositions of 20 advanced Thai learners of English. They found that approximately one-third of all errors were formal, approximately two-thirds were semantic and less than a tenth were attributable to language transfer. Hemchua and Schmitt's framework was chosen as the starting point for the replication study described in this chapter as it built on previous work by James (1998) and Leech (1981), and because its impact is evident in later work (cf. Llach, 2011, Al-Shormani & Al-Sohbani, 2012 and Al-Shormani, 2014a and b). Although several other frameworks have been developed for error analysis (e.g., Dušková, 1969; Zimmerman, 1986a, 1986b and 1987; Meara and English, 1987; Lennon, 1991; Zughoul, 1991; Engber, 1995 and Llach, 2011), Hemchua and Schmitt's is the most comprehensive and recent holistic framework that focuses on LEA for adult language learners.

Using the Hemchua and Schmitt (2006) framework, the compositions of 20 Greek advanced students' compositions were analysed for LEs to establish the number, type, and most frequent error types, and to ascertain how many were attributable to L1 transfer. It was predicted that decisions about the acceptability of lexis would be difficult, as the boundary between what is grammar and what is lexis remains blurred (Lewis, 1993), and that consequently it would be difficult to decide what should and should not be included in the analysis. Problems were also anticipated with allocating error category (type) and cause (what made the learner commit the error?). Particular difficulty was predicted for differentiation between certain types of error in the 2006 framework, for example [Wrong near synonym] and [Collocation] errors).

3.2 Rationale for study

In the context of lexical approaches to ELT gaining ground (Lewis 1993), it is timely to revisit the 2006 study and its applicability to contemporary ELT pedagogies. This chapter replicates Hemchua and Schmitt's (2006) study for the following reasons:

- It would be very useful to establish the reliability and validity of Hemchua and Schmitt's framework. Using compositions from learners of a different L1 in a replication study might provide confirmatory evidence that learners of a similar background at a similar stage of development, but of a different L1, made similar errors in terms of type and number. If results were found to be similar, it would verify Hemchua and Schmitt's (2006:3) claim that their findings would be 'of interest to wider English as a Second Language (ESL)/English as a Foreign Language (EFL) contexts' and would be a first step in investigating whether LE type and frequency are indeed universal across L1 groups.
- The study helps us understand how problematic the issues associated with LEA really are. Previous attempts to conduct EA have encountered problems, such as identification of error, identification of cause of error and classification of error type (Shachter and Celse-Murcia, 1977). It was expected that there would be similar problems in conducting this replication (see Section 3.5 below). Indeed, Hemchua and Schmitt point out that 'in some cases (relatively few), more than one categorisation was possible' (2006:7). However, regardless of the variable precision of categorising LEs, the potential benefits of LEA still make it a valuable exercise.
- The replication can help us find ways to address these problems. Issues in the identification and classification of errors could be identified and addressed to subsequently create an even more user-friendly LEA framework with guidelines that can be used by practising EFL teachers who do not have much training in linguistics.
- The study provides much-needed consistency of evidence, as '[n]o two previous studies on LEs have adopted the same error typology' (Kallkvist, 1998:82). This raises questions of replicability. So there is a growing place for both confirmatory and non-confirmatory replication studies. Porte (2012) argues that although replication studies do not aim for genuine novelty, there are insufficient replications in applied linguistics generally. Porte also claims that such replications are required to establish how second language takes place, and to ascertain whether original findings are reliable and whether they can be generalised to other participants and circumstances.

3.3 Methodology

This section describes the research questions, participants, ethical considerations and how the two studies were conducted.

3.3.1 Research questions

Hemchua and Schmitt (2006) investigated the following research questions using Thai students of English. The replication study used the same questions, but with reference to Greek L1 learners.

- 1) What lexical errors do third-year University students make in their English compositions?
- 2) Which of the errors are the most frequent?
- 3) How many of the errors are attributable to L1 transfer?

3.3.2 Participants

In the 2006 study, there were 20 participants with approximately ten years' English language learning experience (3-5 hours per week). They were in the third year of their undergraduate degree in a university in Bangkok. Their essay brief was 'What are the advantages of country or urban living?' In the 2016 study, participants were 20 Greek students enrolled in a private language school in Athens, studying for IELTS. On average, they had been taught EFL for approximately eight years for two hours per week (less time overall than their Thai counterparts in the original study). Their essay brief was 'Should a government be able to restrict the number of children that a family has?' In both studies participants were similar in age, ranging from 18-26 years old, but factors such as sex and age were not controlled: both groups had little English-writing experience in their primary and secondary schools, but had received some instruction in how to structure an essay, particularly the type of discursive essay found in the data. Both sets of participants were asked to write a 300-350 word argumentative composition without consulting their dictionaries, within 1.5 hours.

3.3.3 Ethics

Participants were told that their writing was being studied, but the focus on vocabulary was not made explicit, as this might alter their performance. Full

informed consent was obtained (see Appendix 3.1) and full ethical approval was gained from the university.

3.3.4 Analysis

The essays were analysed closely following Hemchua and Schmitt (2006): first, the errors were identified and then the correct form of each error was noted by two experienced L1 English teachers. Next, to identify L1 interference, the errors were categorised by a bilingual speaker – Thai and Greek L1 users in the 2006 and 2016 studies respectively. In the first study, the bilingual Thai L1 speaker was the first author: in the replication, an experienced, proficient translator was chosen. These analysed the papers first and consulted with the other authors and experienced EAP L1 English Tutors. The participants were not interviewed retrospectively regarding the meaning behind any of their errors. Allocation of errors to categories (see categories in Table 17 below) was completed with the following rules:

- 1) Erroneous words and collocational phrase errors were included in the count, each counted separately.
- 2) Multiple errors in a phrase were counted separately.
- 3) Exact duplicates of errors in the same paper were counted once.
- 4) When an error could also be classified as L2 or transfer error, it was allocated as a calque error, regardless of the linguistic type (e.g., a [Collocation] error) of error.
- 5) Phrases, such as *You will wake up to *voice's bird*, are classified as a [Connotative meaning] error. (Perhaps Hemchua and Schmitt could have chosen a better example here.)

Grammatical errors were excluded, following these rules:

- 6) In fixed phrases, such as *What's *a matter?*, the error was considered collocational.
- 7) Errors with articles were excluded.
- 8) 'Clause errors' were ignored (e.g., *It's not difficult *for getting to a hospital*). This was interpreted as reduction in adverbial clauses errors (e.g., **While waiting, my hamburger went cold*) and errors in relative clauses (pronoun and referential).

- 9) 'Sentence errors' were ignored (e.g., *I didn't think *how kind they were*). These were interpreted as errors in countability, tense, redundancy, verb agreement and ambiguity/coherence.
- 10) Inter-sentence, or cohesion errors were ignored (e.g., *When someone want's one's help, *he will help each other*)
- 11) Only derivational affix errors (e.g., *He is kind and *considerable*) are included, not plurality, genitive, tense, third person singular, comparative nor superlative.

A Formal Errors	B Semantic Errors
1 Formal misselection 1.1 Suffix type 1.2 Prefix type 1.3 Vowel-based type 1.4 Consonant-based type 1.5 False friends	1 Confusion of sense relations 1.1 General term for specific one 1.2 Overly specific term 1.3 Inappropriate co-hyponyms 1.4 Near synonyms
2 Misformations 2.1 Borrowings 2.2 Coinage 2.3 Calque	2 Collocation errors 2.1 Semantic word selection 2.2 Statistically weighted preferences 2.3 Arbitrary combinations and irreversible binomials 2.4 Preposition partners
3 Distortions 3.1 Omissions 3.2 Overinclusion 3.3 Misselection 3.2 Misordering 3.5 Blending	3 Connotation errors
	4 Stylistic errors 4.1 Verbosity 4.2 Under specification

Table 17 Hemchua and Schmitt's (2006:12) framework for LEA

In Hemchua and Schmitt (2006), both authors analysed the data, but their paper did not discuss rater-reliability. In the 2016 study, two raters (L1 speaker, experienced English Language teachers and examiners) were asked to categorise errors in the first five essays in order to establish whether they could easily use the framework: there was less than 4% disagreement between them (see Section 3.5).

3.4 Results

This section presents the results in terms of comparison of word count, standard deviation, error count, and the types and frequency of errors. In general, the total number of errors and distribution of error types, and therefore the answers to research questions. Number, type of error and the most numerous are remarkably similar to those found in the original study (see Tables 18-20 below). It also

discusses similarities between the two sets of results and offers some implications for these results.

3.4.1 Word count and standard deviation

As can be seen from Table 18 below, the mean length of the 2016 compositions was around 50 words shorter than those in the 2006 study. Also, the earlier study's compositions had a greater range of words than in the 2016 study.

Study	Word count					LEs			
	Total (20 essays)	Mean	Standard Deviation	Min	Max	Total	Mean per paper	Error per number of running words	Percentage
2006	6,906	345.3	81.43	218	578	261	13.1	26.46	3.78
2016	5,912	295.6	47.16	178	407	284	14.2	20.81	4.80

Table 18 Word and error count in the two studies

3.4.2 Error count

Firstly, in the current study, the two experienced L1 English teachers agreed on the identification of virtually every error in their sampling (25% of essays). This contradicts error identification concerns raised by Ellis (1994).

As shown in Table 18 above, there were 261 LEs in the 2006 study with an average of 13.05 errors per paper (one error per 26.46 running words). The 2016 compositions yielded 284 LEs (one error per 20.81 running words).

There were fewer errors in total found in the 2006 study, despite the higher total word count. However, the two total numbers of LEs (a difference of 23 errors) and the percentage of LEs per total word count (a difference of 1.02%) are remarkably similar. Despite the differences in total word count, both the average number of errors per paper and the number of errors per number of running words also showed interesting similarity.

In terms of total word count, there were fewer errors in the Greek essays. There could be several reasons for this. Although both languages have a different script from English, Greek is less distant from English than Thai. There are many borrowings from Greek to English, Greek is an Indo-European language and shows more morphological variation than Thai (word families). Perhaps Greek learners are more attuned to inflexion and derivation. Perhaps this explains why there are proportionally fewer errors in categories A1.1 and A.2. Greek roots are often used to coin new words in English.

3.4.3 Types of errors made

As mentioned, in the 2016 study, two raters (L1 speakers, experienced English language teachers and examiners) were asked to categorise errors in the first five essays in order to establish whether they could easily use the framework for LEA. There was relatively little divergence of opinion as to which category some errors belonged. Where this occurred, problems were noted for discussion (See Section 3.5). Issues were also noted for discussion when the main author categorised the errors. This information would be potentially useful for the development of an improved, future framework for LEA.

3.4.3.1 Formal and semantic errors

As shown in Table 19 below, there was also much similarity in the distribution of the general type of error in terms of formal vs semantic between the two studies: in both, approximately two thirds were semantic and one third was formal. The 2016 study identified slightly fewer formal errors but more semantic errors. The fact that two thirds of the errors were semantic errors underlines the difficulty in semantic knowledge acquisition (sense relation, collocation, connotation and register). Although less frequent, formal errors accounted for approximately one-third of all errors arguing that learners would also strongly benefit from developing their morphological and formal knowledge of lexis ([Misselection], [Misformation] and [Distortion]).

	Formal Errors		Semantic Errors	
	No of Errors	Errors of this type as % of total errors	No of Errors	Errors of this type as % of total errors
2006	96	36.78	165	63.22
2016	82	28.87	202	71.13

Table 19 Summary of frequency in formal and semantic errors

Problems with formal errors

As shown in Table 20 below, the most frequent formal error in the 2006 study was [A1.1 Suffix Type] (9.2% of total errors), underlining the problems that Thai students had with word families. The second most frequent error type was [A2.3 Calque (translation)] errors, but this category only accounted for 6.9% of all errors. This confirms the work by Richards (1971) which states that L1 transfer errors account for only a small portion of total learner errors. In 2016, the most frequent formal

errors were [A2.3 Calque], followed by [A1.1 Suffix Type] and then [A3.1 Omission]. Interestingly, these were the three most frequent categories in the original study also, but the rank order in 2006 was [A1.1 Suffix Type], followed by [A2.3 Calque], then [A3.1 Omission] (see Table 20 below). The totals for [A1.1 Suffix Type] and [A3.1 Omission] were remarkably similar between the two studies, but the number of [A2.3 Calque] errors was almost double in the current study. Perhaps this is due to the fact that the Greek learners knew that their L1 was not very distant from English and they felt that they could use word-by-word translation more confidently to express their intended meanings.

Several categories saw very few errors in either study, confirming that they are lower frequency errors ([A1.2 Prefix type], [A1.3 Vowel-based type], [A1.5 False friends], [A2.1 Borrowing], [A2.2 Coinage], [A3.3 Misselection], [A3.4 Misordering] and [A3.5 Blending]).

Error Type	Rank 2016	Rank 2006	No of Errors (Total=284) 2016	Errors of this type as % of total errors 2016	No of Errors (Total=261) 2006	Errors of this type as % of total errors 2006	No papers containing the error (N=20) 2016	No papers containing the error (N=20) 2006	% of papers containing the error 2016	% of papers containing the error 2006
A1.1 Suffix type	5	3	22	7.75	24	9.20	13	12	65	60
A1.2 Prefix type	10	16	3	1.06	1	0.38	2	1	10	5
A1.3 Vowel-based type	13	17	0	0.00	0	0.00	0	0	0	0
A1.4 Consonant-based type	12	8	1	0.35	12	4.6	1	9	5	45
A1.5 False friends	13	15	0	0.00	3	1.15	0	2	0	10
A2.1 Borrowing (L1 words)	12	17	1	0.35	0	0.00	1	0	5	0
A2.2 Coinage (inventing)	12	17	1	0.35	0	0.00	1	0	5	0
A2.3 Calque (translation)	3	4	34	11.97	18	6.90	15	12	75	60
A3.1 Omission	8	7	12	4.23	14	5.36	10	8	50	40
A3.2 Overinclusion	12	12	1	0.35	6	2.30	1	6	5	30
A3.3 Misselection	12	10	1	0.35	10	3.83	1	7	5	35
A3.4 Misordering	9	13	6	2.11	5	1.92	4	4	20	20
A3.5 Blending	13	15	0	0.00	3	1.15	0	3	0	15
B1.1 General term for specific one	11	9	2	0.70	11	4.21	2	7	10	35
B1.2 Overly specific term	13	17	0	0.00	0	0	0	0	0	0

Error Type	Rank 2016	Rank 2006	No of Errors (Total=	Errors of this type as	No of Errors	Errors of this type as	No papers con-	No papers con-	% of papers con-	% of papers con-
-------------------	------------------	------------------	-----------------------------	-------------------------------	---------------------	-------------------------------	-----------------------	-----------------------	-------------------------	-------------------------

			284) 2016	% of total errors 2016	(Total= 261) 2006	% of total errors 2006	taining the error (N=20) 2016	taining the error (N=20) 2006	taining the error 2016	taining the error 2006
B1.3 Inappropriate co-hyponym	13	15	0	0.00	3	1.15	0	2	0	10
B1.4 Near synonyms	4	1	29	10.21	51	19.54	16	16	80	80
B2.1 Semantic word selection	1	10	89	31.34	10	3.83	18	6	90	30
B2.2 Statistically weighted preferences	13	11	0	0.00	9	3.45	0	5	0	25
B2.3 Arbitrary combinations	13	6	0	0.00	16	6.13	0	10	0	50
B2.4 Preposition partners	2	2	51	17.96	33	12.64	20	15	100	75
B3 Connotative meaning	17	9	0	0	11	4.21	0	7	0	35
B4.1 Verbosity	7	5	13	4.58	17	6.51	9	10	45	50
B4.2 Under specification	6	14	18	6.34	4	1.53	9	4	45	20

Table 20 Rank-order frequency of LEs

Problems with semantic errors

The most frequent error in this area found by Hemchua and Schmitt (2006) was [B1.4 Near synonyms], accounting for 19.54% of all errors. There were three broad error sub-types found in this category: use of informal words for formal ones; non-identical meaning of synonym used and appropriate synonym and two words close in meaning, but different in usage. The second most frequent error in this area was [B2.4 Preposition partners] (12.64%). These could also be sub-divided into three categories (omission of preposition, addition of preposition and substitution of preposition: substitution was the most frequent). These two types were the two most frequent error types overall. However, in the 2016 study, the most common error types were [B2.1 Semantic word selection] and [B2.4 Preposition partners], followed by [B1.4 Near synonyms]. Interestingly, there were over nine times more [B2.1 Semantic word selection] errors in 2016. There were four and a half times more [B4.2 Under specification] errors in the 2016 study (see Section 3.5 for discussion of categorisation issues).

In the two studies, there were very similar numbers (a difference of four or fewer total number of errors) of [B1.2 Overly specific term], [B1.3 Inappropriate co-hyponym] and [B4.1 Verbosity] errors.

3.4.3.2 Most common errors overall

Specific answers to research questions can be found below. The five most common types of errors in the 2016 study were, in order of frequency, [B2.1 Semantic word selection], [B2.4 Preposition partners], [A2.3 Calque (translation)], [B1.4 Near synonyms] and [A1.1 Suffix Type]. In 2006, the five most common error types were, in order of frequency, [B1.4 Near synonyms], [B2.4 Preposition partners], [A1.1 Suffix Type], [A2.3 Calque (translation)] and [B4.1 Verbosity]. In both studies, two categories yielded no errors whatsoever. This shows that errors are indeed not evenly distributed across the error-type spectrum and that some error types are more frequent.

Answers to research questions

1) What lexical errors do Thai/Greek learners make in their English compositions?

Thai learners 2006 - Semantics caused more problems for students than the forms of words. The Thai learners made errors in 21 of the 24 categories above.

The categories where no errors were made were: [A2.1 Borrowing (L1 words)], [B1.2 Overly specific term], [A2.2 Coinage (inventing)]

Greek learners 2016 - Again, semantics caused more problems for students than the forms of words. The Greek learners made errors in all categories, except, [A1.3 Vowel-based type], [A1.5 False friends], [A3.5 Blending], [B1.2 Overly specific term], [B1.3 Inappropriate co-hyponym], [B2.2 Statistically weighted preferences], [B2.3 Arbitrary combinations], [B3 Connotative meaning].

2) Which of the errors are the most frequent?

Thai learners 2006 - The five most common types of errors were, in order of frequency (with percentage of total LEs):

[B1.4 Near synonyms]	20%
[B2.4 Preposition partners]	13%
[A1.1 Suffix Type]	9%
[A2.3 Calque (translation)]	7%
[B4.1 Verbosity]	7%

Greek Learners 2016 – The five most common types of errors were, in order of frequency (with percentage of total LEs):

[B2.1 Semantic word selection]	31%
[B2.4 Preposition partners]	18%
[A2.3 Calque (translation)]	12%
[B1.4 Near synonyms]	10%
[A1.1 Suffix Type]	8%

The 2016 results are very similar to those from 2006.

3) How many of the errors are attributable to L1 transfer?

Thai learners 2006 - The great majority of errors were due to L2 difficulty, not L1 transfer issues. There were 18 [A2.3 Calque] errors, the only interlingual errors to be made, of the three types. This represents almost seven percent of total errors. Greek learners 2016 - Again, the great majority of errors were due to L2 difficulty, not L1 transfer issues. There were 36 (34 [A2.3 Calque] errors, one [A2.1 Borrowing] error and one [A2.2 Coinage] error), accounting for just under 13% of total errors.

In some ways, these are similar findings and support Hemchua and Schmitt's (2006:22) hypothesis that the LEs discussed are likely to be problematic for a wide range of L2 learners. The results are similar in the following ways: the great majority of errors were due to L2 difficulty, not L1 transfer issues, and more interestingly, the same four categories appear in the top five rankings in each study and six of the eight most frequent categories were common to the original study and the replication. Finally, a Spearman's rank correlation coefficient test was run to assess the relationship between the above rankings of the most common categories of the two studies. A two-tailed significance test was also run. A strong correlation was found ($r=0.80$) and results were found to be significant at the 0.01 level. This supports the validity and reliability of the framework.

However, one major finding stood out as quite different from the 2006 study (the number of [B2.1 Semantic word selection] errors: see Section 3.5.3.2 for discussion).

One reason for any differences between the two studies' results is a possible difference in proficiency levels. The precise IELTS or TOEFL scores for the 2016 cohort is unknown, as at the time of data collection they had not yet sat an IELTS or TOEFL test. However, looking at the Greek students' writing, I estimate them to be between IELTS 5.5 and 7. Martin (1984) states that the number of errors found in error analyses does not seem to reduce with higher proficiency levels. However, the higher the proficiency level, the more semantic errors are made.

Only seven types of errors appeared in half or more of the compositions. This shows that learners were making different errors from each other. This in turn demonstrates the potential value of LEA to individual students: if learners could be shown the types and the individual errors that they make, they may take more care in these areas and take steps to reduce them.

3.4.4 Summary of results

- Both studies showed that difficulties with semantics were approximately two times more common than errors in the forms of words.
- There was great similarity between the two studies in terms of frequency of category of errors. Four categories appeared in the top five in each study. They were: [B1.4 Near synonyms], [B2.4 Preposition partners], [A1.1 Suffix Type] and [A2.3 Calque (translation)].

- In both studies, errors were mainly due to L2 difficulty, rather than L1 transfer issues (7% in 2006 and 13% in 2016).
- Only seven categories of errors appeared in half or more of the compositions.
- Several categories of error occurred very infrequently in either study, confirming that errors are not evenly divided across the spectrum. These infrequent types were [A1.2 Prefix type], [A1.3 Vowel-based type], [A1.5 False friends], [A2.1 Borrowing], [A2.2 Coinage], [A3.3 Misselection], [A3.4 Misordering] and [A3.5 Blending].

3.5 Discussion of issues in using the 2006 framework

This section details the issues of error identification, differentiating between grammatical and lexical error, classification and miscellaneous issues.

3.5.1 Problems of error identification

As predicted, it was not always easy to decide whether certain expressions were erroneous, or simply could have been better expressed. However, there were very few instances of disagreement of acceptability between the first rater (the current author) and the two expert EFL teachers and experienced examiners who were asked to analyse the first five essays.

3.5.2 Grammatical vs lexical error

Hemchua and Schmitt (2006) stated some rules (See Section 3.3.4 above) concerning which error types are considered grammatical and should therefore be excluded from the analysis. However, the list proved to be not very comprehensive, some of these rules seemed to lack clarity and were problematic in implementation. For example: 'Clause errors are ignored (e.g. *It's not difficult *for getting to a hospital*)'. This could easily be interpreted as a LE since the word 'difficult' entails a subsequent full infinitive.

3.5.3 Problems of LE classification

Very often, as expected, errors could be assigned to more than one category. This section describes in detail the classification decision issues that were experienced during the LEA.

3.5.3.1 Formal errors

A1.1 Suffix errors

**Before 1980, the *increasing rate was 2.9%, per year*

This could also be categorised as a [B2.1 Collocation] error <The rate of increase was....

*....and *measurements have been taken by the government.*

This could be both a suffix error and the wrong word. (There is no obvious category for wrong word that is not a near synonym.) It could also have been categorised as a [Collocation] error (<...measures have been taken...).

If a learner were to omit a suffix, it could potentially be categorised here or under [B2.1 Semantic word selection].

A1.2 Prefix errors

*.....there are many people who *immigrate in order to.....*

This could also be interpreted as [B2.1 Semantic word selection] error. Perhaps it is both, as 'Wrong word' subsumes a number of more specific error types, so 'wrong word' can (presumably) often be co-classed as another type. It would be better to only allocate 'wrong word' if the entirety of the word is wrong (eg 'I *agree cheese' < 'I like cheese'). The fact that these errors could be allocated to another type may be a problem of analyst consistency, not the categories per se.

A2.1, A2.2 and A2.3 L1 Transfer errors

There is a fundamental flaw with the 2006 framework. Because all L1 transfer errors are assigned to [A 2.1], [A2.2] or [A2.3], it hides the fact that this error may contain, for example an error with a preposition. The confusion here is between type (linguistic type) and cause (L1 interference or complexity of L2). *this is not the only solution *for the problem.* Several calque errors were made with prepositional partners.

However, the Greek L1 speaker and first rater had no problems in identifying erroneous language. As this person is an experienced translator with very good accuracy in English, she was able to identify instances of L1 interference easily. She stated that she did, however, have issues with the number of calque errors in a phrase or sentence. For instance, the sentence below could contain three: **all people there*, **aren't in the globe* and **next a few years*.

However, some of these errors could be seen as grammatical in nature, and not lexical: *This is seen especially in China where *lives half the population of the earth.* Despite the fact that this error is a word for word translation from Greek, the type of error is clearly a syntax error and should therefore be excluded from the study. When analysing language at the phrasal level, problems of allocation to lexical or grammatical error are exacerbated. Also, perhaps this error could be described as one of archaic style (e.g., *Deep in the forest, where live the rat and the mole.*) However, it is less likely that this style was intentionally used.

3.5.3.2 Semantic errors

As expected, this section caused the most difficulty in classification because there were grey areas of acceptability. It was also difficult to decide if the error was [B1.3 Inappropriate co-hyponym], [B1.4 Near synonym] or [B2.1 Semantic word selection]. Associated with this was the difficulty in deciding what phrases occurred in sufficient frequency for them to be classified as collocation errors. This issue was overcome by the author categorising errors to [B2.1 Semantic word selection] if he decided that they were definitely errors within what he felt was a fixed phrase. Perhaps this led to a larger number of [B2.1] errors.

B1.4 Near synonyms

*....and *that is a very serious problem.*

There were instances when it was not clear whether Hemchua and Schmitt (2006) would have classed certain errors as 'grammatical'. For example, errors with deictic pronouns are relatively common, yet their classification is not made clear by Hemchua and Schmitt. However, it is believed that in line with a more modern understanding of what constitutes lexis (Lewis, 1993), these types of errors are more in line with lexical choice and therefore, this quite common error was included in this category.

The same could be said for many/much, a noun countability word choice: *There are *much jobs but there aren't people.*

The same could be said for less/fewer: *Can the government oblige people not to have more or *less children than the government decides?*

*...because every *man has the right to....*

*.....and for *old aged people*

Categorisation of incorrect stylistic choices were included here. For want of another category, 'politically incorrect' language choices, sexist or ageist language could also be considered a stylistic error. However, Category [B4] only had two subcategories: [B4.1 Verbosity] and [B4.2 Under specification].

B2.1 Semantic word selection

The greatest difference between the results from 2006 and 2016 is in the number of errors found in the category, [B2.1 Semantic word selection] (89 in the 2016 study and only 10 in the 2006 study). This could be explained, perhaps, by different categorisation procedures and subjective interpretations of the 2006 guidelines in terms of the differences between [B2.1 Semantic word selection], [B1.4 Near synonyms], [B3 Connotative meaning], etc. In the 2016 study, erroneous semantic word selection was taken as to include selection of completely wrong word, not just in collocations. This was done as there was no other apparently suitable category for this type of error. Perhaps this underlines the need for more detailed guidelines for how to categorise LEs. For example,

*.....by running a project, which is *indicated to limit the number....*

'Indicated' is not a near synonym for 'intended'. 'Which is intended to limit' is not really a collocational phrase. However, without a clear category for wrong word that is not a near synonym, it was included here. Alternatively, more categories may be required to overcome shortcomings such as these with the 2006 framework.

*...there isn't poverty and some people *are very good and *have a good health.*

The two errors in the sentence above illustrate the difficulty in classification between [B1.4 Near synonyms] and [B2.1 Semantic word selection]. The first error could mean 'lucky' and is therefore a [B1.4 Near synonym] error. It could also be construed as a [Collocation] error (<are very well) and therefore be classed as [B2.1 Semantic word selection] error. Similarly, the second error is classified as [Collocational], [B2.1 Semantic word selection], as the correct version should probably be <....are in good health.... or <...are healthy.

*But none of these countries have *released a law like this *as I know.*

The first error is clearly a [B2.1 Semantic word selection] error in the collocational phrase <pass a law. The second could be a collocational error: <...as far as I know. However, it could also be [B4.2 Under specification].

B2.4 Preposition partners

Although prepositions are common, no classification issues were encountered for this category. Hemchua and Schmitt (2006) clearly state that this category should be used for incorrect choice or omission of prepositional partners, as well as inclusion of extra prepositions. It would be useful, however to provide a breakdown of how many of these errors fell into these three sub-categories. Therefore, any improved future framework should take this account.

B4.1 Verbosity

*...breakneck speed *rate*

The inclusion of one extra lexical word could constitute verbosity. If not, a category entitled [Unnecessary word] would be required. This also raises issues of error count: should each extra, unnecessary word count as an error? This simply requires clarification in a guidance document: I suggest that any single or run of unnecessary words should be counted once.

*....scientists will discover at the next centuries planets which will be essential for *the living the people and so maybe the *people solve the problem of population explosion.*

Given the lack of a category for [Cohesion] or [Lexical substitution] errors in the 2006 framework (the student mentions 'people' twice), it is included here.

However, it could also be classified as the wrong word.

B4.2 Under specification

*Although there is the danger of population explosion, it is believed that this is unfair for some people *the moment that it is possible for them to find a way in order to avoid this kind of measurements which are very strict.*

In the 2006 framework, there is no category for [Incoherence]. Therefore, errors of this type were allocated to the [Underspecification] group, as meaning would have been clearer had the learner used more words to convey meaning.

*...the government *not be allowed to limit the number of children a family can have.*

This category could include errors when a single word is missing. However, there

would still be issues with classification: if, for example, a preposition was clearly omitted, it would be a [B2.4 Preposition partners] error.

As can be seen from the issues above, some work could still be done to develop the framework to provide guidance and clarity to the classification process, especially for practising teachers with less training in linguistics.

3.5.3.3 Error Count and Resulting Analysis Issues

In this replication, where the original methodologies for data analysis and reporting were closely followed, there is a lack of clarity in the reporting of types, tokens and repeated tokens. At this point, it would be useful to provide some definitions:

Error types are the different categories of lexical errors that could be made (for example, misspelling (omission), misspelling (addition), etc). Thus, in the sentence, *'The profesor wrote on the blackbord'*, the same type of error (spelling) occurs twice, hence the two errors are counted as one error type.

Error tokens are individual instances of errors of a certain type. In the sentence, *'The profesor wrote on the blackbord'*, there are two occurrences of spelling errors, hence they are counted as two spelling error tokens.

Repeated tokens are where a learner repeats the exact same token more than once in the same script (e.g. *'profesor'* and *'profesor'* are repeated tokens of the same error type).

Hemchua and Schmitt state that repeated error tokens should be considered as one error type and counted once. Some may view this as sensible in that counting all repeated tokens would make a piece writing look more erroneous than it really is if one considers there may be only a few error types but repeated several times. However, a balanced view would be that ignoring repeated tokens would not present a full picture of a learner's work. Omission of repeated tokens also causes other issues: although Hemchua and Schmitt's (2006) methodology was followed precisely in that repeated tokens were omitted from both studies, the way the findings are presented may disadvantage those students using a wider repertoire of lexis because if a learner uses a greater range of lexical types, they may produce more error tokens. On the same principle, a learner restricting themselves to repeated lexical items may have fewer errors counted. Furthermore, the data in Table 19 above states the total number of tokens (excluding repeated tokens) as a percentage of the total number of words. However, this cannot be accurate if repeated tokens are omitted from the error counts. It is also problematic to compare

the total number of errors in Table 18, when one study has 6,906 words and another has 5,912. On reflection, one could have made the two word counts the same, but then the number of essays would have been different.

A better way of comparing these data in the current study, when following the original methodology for the sake of accurate replication, would be to include a focus on the error types made. Therefore, in the 2006 study, 10 formal error and 10 semantic error types were found, whereas in the 2016 study, there were 10 and 6, respectively. As can be seen, there are similarities, but this is a rather surface level analysis of results. See Table 20 above for further discussion.

There seems to be a cyclical issue here in that counting types only in LEA ignores tokens, and counting tokens only ignores types. Reporting error types provides a picture of the breadth of lexical errors made, an interesting insight into which areas of lexical development can be improved, whereas reporting tokens illuminates which types learners are struggling with the most and may provide focus for teaching. Given the issues above, it would be preferable to include repeated tokens in the error count and present the number of types and tokens when conducting future LEAs.

3.5.3.4 Miscellaneous Issues

As well as clarification of the issues above, there are some other areas for improvement:

- There is no category for when one word was incorrectly written as two (e.g., **every one*).
- The framework lacks a category for language that is not necessarily wrong, but awkward in expression. This would help with trying to decide on acceptability.
- The framework would be improved with the introduction of a category for inappropriate slang or poor lexical selection for genre.
- The lack of an [Incoherence] category needs to be addressed. When a learner produces an incoherent statement, it is difficult to categorise the error due to the fact that a plausible interpretation of the error cannot be made. This may lead to the error not being categorised or allocated randomly to a category. The result of this would be that these errors would not be

highlighted. Since incoherence errors are more serious in that they cause breakdowns in communication, it is important that they are dealt with systematically. Despite this being an apparent move away from the specificity of the framework, the addition of an [Incoherence] category would be very beneficial to the learner.

- A final addition would be the inclusion of a [Lexical cohesion] error category. (...scientists will discover at the next centuries planets which will be essential for **the living the people and so maybe the *people solve the problem of population explosion*).

These errors could be considered lexical and were quite numerous in the Greek data, and would be quite simple to remedy if they are given attention.

3.6 Implication for practice

The results have implications for the teaching of lexis: if these issues are indeed universal for all L1 groups studying English as a Foreign or Second Language, a greater focus on collocation and word families is required. This could be done by providing contextualised, authentic input, ensuring the noticing of collocational partners and by encouraging learners to use corpus linguistics to investigate collocations. Dictionary work and the completion of 'word family trees' would help to familiarise learners with different word family members.

This replication has provided some confirmatory evidence to support Hemchua and Schmitt's (2006) hypothesis that learners of a similar background at a similar stage of development, but of a different L1, may make similar LEs in terms of type and number. This verifies their (*ibid* 2006:3) claim that these findings would be 'of interest to wider English as a Second Language (ESL)/English as a Foreign Language (EFL) contexts'. Hopefully, this study will help to fill a gap in LEA research, and re-vitalise interest in LEA by encouraging practicing teachers to conduct LEAs of their own.

3.7 Conclusion

The results of the current study were remarkably similar to those found in the original, despite the fact that gender, age and proficiency level were not controlled, and a different L1 of students and a different essay brief was used. This suggests that the 2006 framework is relatively fit for purpose, despite the issues discussed in Section 3.5, and that it can be used with different L1 groups. There were, however, some issues found during the LEA implementation, namely, difficulties in deciding

whether a clause contained an error, whether that error should be considered grammatical or lexical, and if the latter, to which category it should belong. These problems were expected. There were also issues with not counting repeated tokens and the effect this would have on the way the data are reported. I feel these should be included in future LEA. More specific issues of categorisation when using the framework include: whether a LE should be considered connotative or collocational; whether errors should be allocated to just [Calque] or to another category also; whether an error occurred in a fixed expression (or whether that expression is indeed a fixed expression or not) and should be allocated to [Semantic word selection] or whether it should be allocated to another category; lack of clarity over where to allocate single extra or omitted words and also, finally, where to allocate a completely incorrect word choice that was not a [Near synonym] or [Inappropriate co-hyponym].

Results would have been more comparable if the participants had been of exactly the same L1, age, gender split, and proficiency level, and the same essay brief had been used. Unfortunately, I did not have access to such a group, although the difference in variables allowed for some comparison across these variables.

Some of the guidance for error identification was unclear in the 2006 publication, which may have led to inconsistent analysis. I could have contacted the original author to check I was conducting the LEA as intended, perhaps ensuring closer results.

Due to the fact that much similarity was found between the type and frequency of LEs between Thai and Greek learners of English, further research into the errors made by students of other L1 groups would help establish whether similarity is as widespread as suspected by Hemchua and Schmitt (2006) and the current author. If this is indeed found to be the case, it would strengthen the call for the teaching recommendations in 3.6 above. Unfortunately, there does not appear to be much current research into this area.

Work should also be conducted to create an improved framework to analyse LE. Suggestions include:

- Clearer guidelines on what to exclude in terms of grammatical items
- Including all repeated errors in the count

- Clearer guidelines on allocation of LEs to category when there is more than one possibility
- Separation of cause and type of error
- Further sub-categories that will allow for categorisation of all errors ([Coherence], [Cohesion], [Awkward expression], [Missing word], and more precise allocation of sub-types of errors. (i.e., whether some error categories, e.g., a [B2.4 Preposition partners] error is an [Omission], [Addition] or [Substitution] error, for example).

The next chapter will examine further frameworks for conducting LEA. They will be tested using selected essays from the same Greek data that were used in this study. The advantages and disadvantages of these frameworks, along with the issues uncovered in this study, will be used to inform the design of a new LEA framework (Chapter 5). After that, the revised framework is tested on further compositions by L2 learners.

Chapter 4 (Study 2) Learning from Existing Lexical Error Analysis Frameworks: A Comparative Approach

4.1 Introduction

The previous chapter described Study 1, a replication study of LEs using Hemchua and Schmitt (2006). It was found that there was some similarity in results in terms of the most common error types between the two cohorts (Thai and Greek), suggesting that learners of different L1 groups did indeed make similar LEs. This suggested in turn that there was value in LEA. However, Study 1 included the use of one framework only. Despite the similarity of results, the Hemchua and Schmitt (2006) framework is still problematic in that it was possible to categorise various LEs into more than one area. It seems that a more straightforward framework is still required if LEA is to become a practical tool for practising teachers or SLA researchers. The present chapter is an investigation into the advantages and disadvantages of a wider range of previously published frameworks that examine LEs, intended to provide data for the creation of a more user-friendly and useful framework.

Compared with grammatical or whole language EA, which attracted some criticism in the past (e.g., Schachter and Celce-Murcia, 1977) there have been relatively few studies that focus on LEs alone (Hemchua and Schmitt, 2006). Given the importance and centrality of lexis in language learning and teaching (see Section 2.5.1), it would be useful if there was a standardised, user-friendly framework for analysing learners' LEs. Such a tool could be used by researchers for investigating lexical acquisition and by teachers for collecting information on and pointing out LEs to individual learners. It could also inform remedial whole class teaching. Since the late 1960s, a small number of frameworks for LEA have been published: Dušková (1969), Richards (1971), Corder (1973), Zimmerman (1986), Meara and English (1987), Zimmerman (1987), Lennon (1991), Zughoul (1991), Engber (1995), James (1998), Hemchua and Schmitt (2006), and Llach (2011). These have been quite varied in terms of investigation of the type and cause of LE, resulting in mixed results and problems with error identification and categorisation. These frameworks seem to vary in terms of ease of use (with respect to categorisation of all errors and propensity for allocation of error into more than one category) and the depth of analysis they produce (detail of results). It seems possible that the more categories there are in a framework, the greater the

possibility of allocation to more than one category, and the fewer categories there are, the less depth of analysis the framework can provide. However, the main issues with LEA seem to be identification of error and the confident allocation of errors to categories. Therefore, the matter of clear delineation of category is problematic and of particular importance. Lott (1985:259) highlights the difficulty in building a system of definitions where 'the analyst will not periodically have doubts about how to categorise particular errors: the quality of any research must be affected by the researcher's intuitions about [the type of error and] why the error occurred.' It is hoped that by removing as much ambiguity as possible, an optimum balance between ease of use and depth of analysis can be found in a new, improved framework. The literature review (Chapter 2) has described the main previous LEA frameworks taken from published studies. There was some discussion of their potential ease of use and potential depth of analysis, with an examination of the potential benefits and drawbacks of their implementation and use. To further test their ease of use and depth of analysis, six previously published frameworks for EA will be utilised in this study.

To ensure a useful degree of comparison, this chapter uses the same Greek essay data set that was used in Chapter 3. The five most erroneous essays were selected from the replication in Study 1 and analysed using the frameworks above. The frameworks were tested for the following criteria: ease of use (distinctiveness of categories, the possibility of allocation to more than one category of error and the number of categorisable and uncategorisable LEs), and depth of analysis (appropriateness, spectrum and number of categories).

The advantages and disadvantages of these frameworks were analysed to create a more effective framework for investigating LEA (See Chapter 5).

4.2 Methodology

The main aim of the study was to ascertain the advantages and disadvantages of selected LEA frameworks so that an improved framework could be created. Six frameworks were selected as they represented various approaches to LEA, employed different types of frameworks and were often informed by linguistic research and variety in categorisation or description of lexis. These frameworks are: Dušková (1969), Corder (1973), Zimmerman (1986), Engber (1995), James (1998) and Hemchua and Schmitt (2006). The latter was used in Study 1. To explore the advantages and disadvantages of these varied frameworks and to

extract pointers to develop an improved framework, the 'ease of use' and 'depth of analysis' of these frameworks had to be established. 'Ease of use' refers to distinctiveness of categories or how easy it was to allocate errors to categories: this is important when one considers that some English teachers around the world are not experts in linguistics and may struggle to use some of the frameworks that use more specialist linguistic terminology. The section also reports on how much I had to deliberate over which category to allocate to. The number of uncategorisable errors (errors that do not fall into the framework categories) is established. 'Depth of analysis' refers to the number and appropriateness of categories, i.e., the extent to which the final analysis shows a sufficiently wide range of error types (i.e. more than 10 types) to make it useful to both the SLA researcher and the teacher/learner seeking to benefit from its use. Although only six frameworks were used in the study, the potential benefits of six further frameworks (Richards (1971), Zimmerman (1987), Meara and English (1987) Lennon (1991), Zughoul (1991) and Llach (2011), are gauged. These were also included below. In order to investigate these areas of the various LEA frameworks, five essays were selected for their number and range of LE type (see Section 4.2.4.1).

4.2.1 Research questions

The following research questions form the main focus of the study:

1. How easy to use is each framework? (What is the proportion of uncategorisable errors, and to what extent is there the possibility of allocating errors to more than one category?)
2. What depth of analysis does it provide? (Is there a sufficient range of error types exposed by the framework, making it useful to the SLA researcher or teacher/learner?)
3. What can be taken from the frameworks and applied to a new improved framework? (What categories should be used? What changes might be necessary?)

4.2.2 Participants

Five of the same essays that were used in Study 1 were re-used here. See Section 3.3.2 for a description of the participants and their essay data.

4.2.3 Ethics

See Section 3.3.3 for details of ethical procedures.

4.2.4 Analysis

This section describes how the data were selected and analysed.

4.2.4.1 Data selection criteria

Five of the 20 participants' essays were chosen. In order to thoroughly test the frameworks, the essays had to conform to four out of five of the following criteria, established through LEA using the Hemchua and Schmitt (2006) framework in Study 1. Four of the essays (excluding No13: no uncategorisable error) conformed to all five criteria below.

- 1) From the 20 essays, the essay contains the most LEs.
- 2) The categorisation of the majority of these is not entirely straightforward using the Hemchua and Schmitt framework. (I.e., there is a large number of errors that could have been categorised in more than one area.)
- 3) The essay contains at least six different types of LE.
- 4) These errors include transfer and non-transfer errors.
- 5) The essays contain at least one error that was uncategorisable using the Hemchua and Schmitt (2006) framework.

4.2.4.2 Process

Firstly, all the L1 errors in the five essays were identified and categorised into Calque, Borrowings and Coinage Errors by a Greek L1 speaker and professional translator who had excellent English skills. Secondly, for consistency, grammatical errors were identified so that they could be excluded from the study. This was done following guidance provided by Hemchua and Schmitt (2006), as described in Section 3.3.4.

Next, LEA was carried out on the five essays using each of the six frameworks: the LEs were identified, categorised and counted. The total numbers of categorisable and uncategorisable errors were noted for each essay, as well as their representative percentage totals for comparative purposes. Examples of dual categorisation possibilities were also noted for discussion. Hemchua and Schmitt's

(2006) procedure for counting LEs was followed, as described in Section 3.3.4. However, all repeated tokens were included in the analyses.

It was decided not to interview the students regarding the meaning behind any of their errors (as suggested by Corder, 1973), as this was an extra, time-consuming, unscientific step in LEA which may not always be possible and therefore should not be included in the process of LEA. A framework was sought that did not require this stage.

4.3 Results and discussion

This section will present the results of the LEAs and discuss main findings in terms of the most frequent LE types, ease of use (distinctiveness of categories or degree of confusion in terms of allocation of error to more than one category and number of categorisable vs uncategorisable errors) and depth of analysis (appropriateness of categories, number of categories). Suggestions from each analysis for a new, improved framework will also be presented. For a more detailed description of these frameworks, see Section 2.5.7.

4.3.1 Dušková (1969) (Etiologic/process-oriented)

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
1 Formal similarity	0	1	0	0	0	1	0.8
2 Relatedness of meaning	12	6	2	11	0	31	25
3 Assumed equivalence	4	2	2	2	2	12	9.7
4 Distortion	3	0	1	1	4	9	7.25
Total number of errors identified by the framework	19	9	5	14	6	53	42.8
Other errors							
preposition partner errors	2	7	9	2	4	24	19.5
stylistic verbosity/ unnecessary word	1	1	0	0	2	4	3.2
completely wrong word	0	4	3	2	8	17	13.7
when two words should be one	0	0	0	2	0	2	1.6
miscellaneous errors with coherence of clauses/ phrases.	5	0	1	2	1	9	7.3
stylistic underspecification/ missing word (s)	6	1	4	1	3	15	12.1
Uncategorisable total	14	13	17	9	18	71	57.3
Total number of errors	33	22	22	23	24	124	100%

Table 21 Number of errors in each category and the number of uncategorisable errors using Dušková (1969)

4.3.1.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 53 error tokens were found across four error types. Interestingly, only one formal error (Category 1) was identified in the five essays, accounting for just 0.8% of all errors. The highest number of errors, 31, fell into Category 2, [Relatedness of meaning], accounting for 25% of all errors. The most remarkable finding was the large number of errors that were uncategorisable using Dušková's framework (57.3%). Of all these uncategorisable errors, [Prepositional partners was the largest] (19.5%), followed by [Completely wrong word] (13.7%).

4.3.1.2 Ease of use

It was not possible to allocate the majority of the 124 LEs to categories. This has serious implications when using the framework. The framework was quite simple to implement and could be useful for teachers with little linguistic training or understanding of the wide variety of LEs that could be made. It was relatively easy to spot errors of formal similarity, (imprecise synonyms and spelling errors or words that do not exist in the English Language. (Categories 1, 2 and 4 respectively). Grouping all spelling errors (omission, insertion and misselection) appears to be an efficient way of dealing with spelling errors, as separation serves a limited purpose. There were very few instances (3) where error could have been allocated to more than one category, which is unsurprising, given the number of categories. For example, the [Formal similarity] error (Category 1) error, **...see other solution...* could have been categorised as a [Relatedness of meaning] error" (Category 2). I made the decision to put all dependent preposition errors into the uncategorisable group. Perhaps they could also have gone into the [Relatedness of meaning] category (C2), but Dušková does not comment on this. It is possible that the Category 2 errors [Relatedness of meaning] could encompass other error types. This is a case for excluding cause of error from initial LEA and focussing only on type, as being certain of the cause is not always possible from looking at the error. A focus on cause could come at a later stage in error analysis, if possible or required.

4.3.1.3 Depth of analysis

Having only four categories for LE categorisation makes presentation of errors very straightforward for learners and teachers. However, this is also the main flaw with the framework: it does not give a detailed picture of the types of errors made. A further interesting advantage is that each of the four categories of error relate to

the causes behind them, which may help learners to understand why they were made. However, it is plausible that when their errors are pointed out, learners may be able to see for themselves which may have been committed due to L1 interference. Category 3, [Assumed Equivalence], obviously requires the eye of a skilled bi-lingual analyst. Assuming such a person was available, there is still a lack of sub-categories such as [Borrowings] and [Coinage]. Further, the original examples in Dušková (1969) were of the single word variety, with no mention of whole phrases translated from L1. Using the framework with individual lexemes would result in error count issues because when a phrase is made up of questionably selected lexical items, it is difficult to point to individual words as errors when, in fact, the whole phrase could be better counted as one error. There were no categories for the following types of error: [Preposition partner] errors (however, Dušková may have intended for these to be categorised under 2), [Unnecessary word/verbosity], [Underspecification/missing word] or [Miscellaneous errors with coherence of clauses/phrases]. Of these, perhaps the most serious omission was errors of coherence. These are, in a sense, the most serious of error types as they cause breakdowns in communication.

Overall, the large proportion of uncategorisable errors and restricted number of categories makes this an unsuitable framework for detailed LEA.

4.3.1.4 Points to consider when creating a new framework for LEA

- Exclude cause of error from LEA and focus only on type to avoid dual categorisation
- Retain grouping of spelling errors under one category of [Distortions].
- Include a category for [Preposition partner] errors and sub-categories for differentiation between [Omission], [Addition] and [Misselection].
- Include a category for [Completely wrong word].
- Include a category for [Two words should be one].
- Include a category for [Miscellaneous errors with coherence of clauses/phrases].
- Include a category for [Stylistic underspecification/missing word].
- Include a category for [Unnecessary word/ stylistic verbosity].
- If investigating cause of error, include categories for [word-by-word translations of clause or parts of a clause], [Borrowings] and [Coinage].

4.3.2 Corder 1973 (Grammatical or linguistic criterion)

Category	P 4	P 5	P 6	P 12	P 13	Total No Errors	% of all errors
1 Omission	3	4	3	2	4	16	13
2 Addition	2	3	1	1	3	10	8
3 Selection	21	16	15	14	14	80	66
4 Misordering	1	0	0	0	4	5	4
Total number of errors identified by the framework	27	23	19	17	25	111	92
Other errors							
Miscellaneous	0	0	0	0	1	1	.8
When two words should be one	0	0	0	2	0	2	1.6
Miscellaneous errors with coherence of clauses/ phrases.	3	0	1	3	0	7	6
Uncategorisable total	3	0	1	5	1	10	8
Total number of errors	30	23	20	22	26	121	100%

Table 21 Number of errors in each category and the number of uncategorisable errors using Corder (1973)

4.3.2.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 111 error tokens were found across four error types. As can be seen in Table 22, the most frequent error type was that of [Selection]. 80 errors were found, accounting for 66% of all LEs. The next most frequent type was [Omission]. 16 errors were found in this category, accounting for 13% of all errors. This was followed by 10 [Addition] errors (8%) and finally, five [Misordering] errors were found (4%). What is remarkable is that there are over five times as many of the most frequent errors as there are of the second most frequent type. Perhaps this can be explained by the catch-all function of a general [Selection] category. Due to the broad nature of these categories, only 8% of the 120 errors identified were uncategorisable. The majority of these were [Coherence] errors.

4.3.2.2 Ease of use

This is probably one of the easiest frameworks to use because of the limited number of categories. However, distinction between grammatical and lexical error would be easier had Corder defined or delimited the boundaries of each area or offered detailed guidance in differentiation. Corder's framework does not seek to attribute cause at this stage in analysis, making the framework even easier to use.

Avoidance of speculation of cause of error made the process faster and avoided the speculative issues of categorisation under transfer error or developmental type of error. It was particularly easy to allocate errors to the [Selection] category, as this served as a catch-all for incorrect prepositional partners and individual words and parts of a phrase, even when the original intended meaning of that clause was unclear, hence the much smaller number of [Coherence] errors.

There were very few instances where errors could have been placed in more than one category. In four instances there was deliberation over whether a phrase that contains the wrong words [Selection] could also have been placed in the [Omission] category, because that phrase contained too few words to adequately state the desired meaning, or in the [Addition] category because it contained too many.

This simple framework would be useful for teachers with limited linguistic training, for teachers wishing to offer feedback to students and for teachers of multilingual groups who are not familiar with all the students' L1s. It is not difficult to comprehend the concept of missing word(s), extra word(s), wrong word (s) or wrong word/letter order, but the decisions when categorising may be difficult. Overall, categorisation of error was relatively straightforward with only five categories (including [Uncategorisable]). There were only 10 instances (8% of all errors) where errors were uncategorisable under the framework. Given this low percentage of uncategorisable errors, the broad framework of [Omission], [Addition], [Selection], [Ordering] seems to capture the vast majority of errors, and perhaps subcategorisation of these broad areas would suit an improved framework well, as it may retain simplicity.

4.3.2.3 Depth of analysis

As can be seen from Table 22 above, the vast majority of errors were allocated to the [Selection] category. This does not allow the analyst to easily see the range of errors, which could be viewed if a more extensive taxonomy had been used. It also hides [Coherence] errors, which should be dealt with very carefully, if we consider communication breakdown as the more serious type of error. Further categorisation is required if one would like to know what type of LEs were omitted/added or misselected (e.g., [Spelling], [Near synonym], [Prepositional partner], etc, etc). Perhaps the trade-off between ease of use and depth of analysis is too much in favour of the former here. Further, the framework should

allow for the separation of [Omission], [Addition], [Selection], [Ordering] of letters in a word [Spelling] and words in a clause. It did not allow for categorisation of errors in the following areas: [Connotation], [Collocation], [Prefix/suffix], [Lexical cohesion] and [Coherence]: it is this level of detail that LEA should expose.

4.3.2.4 Points to consider when creating a new framework for LEA

Include a range of categories which will allow for analysis of the full spectrum of LEs. It would be valuable to retain the [Omission], [Addition], [Selection], [Ordering] distinction, as this seems to account for the vast majority of errors. This taxonomy could be applied to a greater variety of types of LEs (as mentioned in the previous section), so one could have, for example, [Misselection of prepositional partner], etc.

4.3.3 Zimmerman (1986) (Descriptive/product-oriented)

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
1) Sense Relations Errors	3	0	2	8	2	15	12
2) Field Errors	4	4	3	4	5	20	16
3) Feature Errors	0	0	1	0	0	1	1
4) Word-formation Errors	7	2	1	1	4	15	12
5) Collocation Errors	4	6	7	6	2	25	20
6) Idiomatic Expression Errors	0	0	0	0	0	0	0
7) Omission Errors	5	2	3	0	3	13	10
8) Redundancy Errors	2	2	0	0	3	7	6
9) Paraphrase Errors	8	2	1	5	3	19	15
10) Stylistic Errors	0	1	0	0	0	1	1
11) Connotative Errors	0	0	0	0	0	0	0
Total number of errors identified by the framework	33	19	18	24	22	116	91
Other errors							
When two words should be one	0	0	0	2	0	2	2
Miscellaneous errors with coherence of clauses/ phrases.	6	0	1	1	1	9	7
Uncategorisable total	6	0	1	3	1	11	9
Total number of errors	39	19	20	26	23	127	100%

Table 22 Number of errors in each category and the number of uncategorisable errors using Zimmerman (1986)

4.3.3.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 116 error tokens were found across 11 error types. The most common error type with this framework was that of [Collocation]. 25 errors of this kind were found, accounting for 20% of all errors. The next most common type was [Field] errors. 20 of these were found, accounting for 16% of all errors. This was closely followed by 19 [Paraphrase] errors (15% of all LEs). The latter is unsurprising, given the nature of lexis as chunks and the associated difficulties of pinpointing covert errors in a phrase. Only eleven uncategorisable errors were found.

4.3.3.2 Ease of use

Avoiding speculation as to the cause of the error allowed for faster analysis. Allocation to category was quite straightforward for many errors, but very often dual categorisation was possible, as predicted. For instance, errors in Categories 5 and 10 could also be categorised under 9. As can be seen from Table 23, there were relatively few uncategorisable errors with this framework (9%). The majority of these were [Phrasal errors of coherence]. However, many other errors were phrase-based and did not fit into the somewhat out-dated view of lexis as single words. Some of these would have been categorisable if there had been columns for determiners, conjunctions, prepositions and pronouns. It could be that Zimmerman considered these as grammar words and therefore excluded them from his original framework. [Phrasal] errors were still included under the 11 categories, in a fifth column. The distinction between [Sense relations] and [Field] and [Word formation] error was quite straightforward when dealing with individual words. It was, on a few occasions, difficult to differentiate between [Sense relations] and [Feature] errors. Zimmermann himself reported problems with decisions in these areas. It will always be difficult to decide what constitutes a collocation or formulaic chunk. For example, **In all the world* (all around the world) could be classified as a [Collocation] or [Paraphrase] error. Clarification is required as to whether the error should be counted under the part of speech that appeared as an error or the part of speech that it should have been: the former was chosen in this study to avoid dual categorisation issues, which may be due to speculation of what the intended word should have been. It was not clear whether some phrase errors could be classified under [Redundancy]. Further guidance was required for use. For example, if [Phrasal] errors could not be easily categorised under other areas, then they should be placed in [Paraphrase] error.

4.3.3.3 Depth of analysis

The number of categories provides for a broad spectrum of errors and therefore good depth of analysis. Zimmerman provides several categories for when the wrong word is used, again allowing for more depth of analysis. It is not clear why Zimmermann chose to analyse errors in terms of word class. It adds depth, but seems to be of limited value to learners and teachers. Category 5 provided a place for the large number of prepositional partner errors, but a separate category for these would provide a clearer picture of the type of errors being made. As mentioned above, the [Collocation] category proved to be the most frequent, but this is of limited use to the analyst who wishes to look at specific types of errors. The analyst may be tempted to allocate incoherent phrases to the [Paraphrase] group, but this would hide the more serious coherence errors, so a [Coherence] category is called for.

4.3.3.4 Points to consider when creating a new framework for LEA

- Add a new category entitled [Coherence].
- Provide guidance for dual categorisation possibilities. A hierarchy based on seriousness of error would be useful.
- Retain categories for [Sense Relations], [Field] and [Word Formation] errors with subcategories for [Spelling] and [Two words should be one].
- Include a separate category for [Prepositional partner] errors.
- Retain a category for [Feature] error, but clarify the difference between feature (verbose description of a lexeme that includes one, some or all features of that lexeme) and [Paraphrase] error (a miscellaneous error that is deviant and requires different words, perhaps in a different order, to convey intended meaning).
- Analysis of part of speech allows for more depth, but the information may be of limited use to learners and teachers. Perhaps the horizontal axis could be replaced by Corder's [Omission], [Addition], [Selection], [Ordering]. This would offer more useful remedial information and enable the new framework to retain all categories except 7 and 8.

4.3.4 Engber (1995) (Etiologic/product-oriented)

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
1A1) Incorrect word choice– semantically unrelated	5	3	3	1	5	17	13
1A2) Incorrect word choice – semantically close	2	2	3	8	7	22	17
1B1) Two lexical items	0	1	1	0	0	3	2
1B2) Phrases	5	4	6	8	7	30	23
1B3) Multiple errors involving core lexical items	6	0	1	1	1	9	7
2.1) Derivational errors	2	0	0	1	0	3	2
2.2) Verb forms	0	3	0	3	1	7	5
2.3) Phonetically similar, semantically unrelated	0	0	0	0	0	0	0
2.4) Word distorted-major spelling error	2	0	1	1	4	8	6
Total number of errors identified by the framework	22	13	15	23	25	99	75
Other errors							
Missing Word	4	2	2	0	3	10	8
Incoherence	8	0	1	0	0	9	7
Unrequired word	1	1	0	0	3	5	4
Verbosity stylistic	0	2	1	1	0	4	3
Repetition	0	1	0	2	0	3	2
Two words should be one	0	0	0	2	0	2	2
Uncategorisable total	13	6	4	5	6	34	26
Total number of errors	35	19	19	28	31	132	100%

Table 23 Number of errors in each category and the number of uncategorisable errors using Engber (1995)

4.3.4.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 99 error tokens were found across 9 error types. The three most frequent types of error using this framework were [Phrases] (30 errors, accounting for 23%), then [Incorrect word choice – semantically close] (22 errors accounting for 17% of all errors) and then [Incorrect word choice– semantically unrelated] (17 errors accounting for 13%). Again, the distinction between the latter two proved useful in separating a potentially single [Wrong word] category. There was a large gap between the third most common category and the remainder with three categories yielding fewer than four errors each. There were 34 uncategorisable errors with a fairly even distribution across the spectrum of error types – a relatively high number, given the total number of categories. This suggests that it is the type of category and not

the number of categories that influences the number of uncategorisable errors in a framework's application.

4.3.4.2 Ease of use

With relatively few categories (nine), allocation of error should have been less problematic, but there was some deliberation, and the framework did not allow for the categorisation of 25% of LEs. In fact, with no further explanation of categories or guidance, it is difficult to imagine a problem-free classification using this system. However, without seeking to attribute cause to the various errors, such as L1 interference, the framework is made easier to use. Category [1A2] groups sense relations, rather than separating them, making for quicker analysis. [1B2] provides a [Paraphrase] category avoiding the necessity to 'unpick' multiple errors in phrases. Including a [Verb form] category is a good idea as it allows for a grammaticised lexis view.

The following problems were encountered. Decisions between whether an incorrect word choice is semantically related can be a matter of interpretation. It is unclear how useful it is to differentiate between errors made with two words or a phrase. This could also result in dual classification, e.g., **In the other side*. It would also be possible to categorise some lexical form errors in both categories. It is very useful having a category for errors with core lexical items (Bell 2012). However, focussing learners' attention on errors with high frequency lexical items makes sense, but this may make the LEA process longer as the analyst will need to be fully familiar with the list of core items. It was not clear how distorted the spelling has to be before it becomes [Semantically unrelated] nor what constitutes a 'major' spelling error. Engber mixes descriptive categories [Spelling] errors and interpretive ones [Confusion due to phonological similarity]. It may be a subjective choice between the two. The category [Phonetically similar, semantically unrelated] could cause categorisation confusion with [Spelling]. Further, it is not always easy to decide what a [Phrase] is. Sometimes it was difficult to decide between [Multiple errors involving core lexical items] and [Phrase] errors, for example, **In the all world*. There were no categories for [Missing word], [Incoherence], [Addition], [Verbosity] (stylistic or repetition) or [Two words should be one].

4.3.4.3 Depth of analysis

With only these nine categories, a clear picture of the type of LEs being made could not be gained. Without other categories, the teacher would be quite limited in terms of suggestions for remedial work or which aspects of word knowledge to teach in the future. Conspicuous by their absence are [Verbosity], [Underspecification], [Incoherence], [Cohesion], [Style] and [Preposition] error. [1B2] provides a place for errors with prepositional partners. However, the presence of a sub-category for these would help to paint a more detailed picture. [1B1] describes the number of errors in a clause, not type. The category could include an [Omission] and an [Incorrect word] choice.

4.3.4.4 Points to consider when creating a new framework for LEA

- Provide explanation of categories or guidance for categorisation
- Include a [Missing word] category
- Include an [Addition] category
- Include an [Incoherence] category
- Include a [Verbosity/stylistic/repetition] category
- Only have one category for [Two lexical items] and [Phrase] error.
- Avoid causal categories, such as [Phonetically similar, semantically unrelated].

4.3.5 James (1998) (form- and content-oriented)

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
Spelling Omission	2	1	1	1	1	6	5
Spelling Over-inclusion	0	0	0	0	0	0	0
Spelling Misselection	0	0	0	0	0	0	0
Spelling Misorder	1	0	0	0	3	4	3
Spelling Blend	0	0	0	0	0	0	0
RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc. Omission	4	2	3	1	3	13	10
RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc. Over-inclusion	0	2	0	0	3	5	4
RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc. Misselection	11	10	9	11	12	53	41
RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc. Misorder	0	0	0	0	1	1	1
RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc. Blend	0	0	0	0	0	0	0
Sense relations/ collocations omission	0	0	0	0	0	0	0
Sense relations/ collocations Over-inclusion	1	0	0	0	0	1	1
Sense relations/ collocations misselection	2	3	3	3	1	12	9
Sense relations/ collocations misorder	0	0	0	0	0	0	0
Sense relations/ collocations blend	0	0	0	0	0	0	0
Cohesion/ coherence/ genre-fidelity/ felicity omission	0	0	0	0	0	0	0
Cohesion/ coherence/ genre-fidelity/ felicity Over-inclusion	0	0	0	0	0	0	0

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
----------	----	----	----	-----	-----	-----------------	-----------------

Cohesion/ coherence/ genre-fidelity/ felicity misselection	12	2	3	2	2	21	17
Cohesion/ coherence/ genre-fidelity/ felicity disorder	0	0	0	0	0	0	0
Cohesion/ coherence/ genre-fidelity/ felicity blend	0	0	0	0	0	0	0
Total No Errors recognized by the framework	33	20	19	18	26	116	91
Other errors							
Paraphrase errors	2	1	2	3	1	9	7
Two words should be one	0	0	0	2	0	2	2
Uncategorisable total	2	1	2	5	1	11	9
Total number of errors	35	21	21	23	27	127	100%

Table 24 Number of errors in each category and the number of uncategorisable errors using James (1998)

4.3.5.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 116 error tokens were found across 20 error types. The most common error type here was clearly [Misselection of RANK: Clause-Phrase-Word-Morpheme/ CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc]. (53 errors accounting for 41% of all errors). Many of these were made up of errors with prepositions. The next most common error type was [Misselection cohesion/ coherence/ genre-fidelity/ felicity] with 21 errors, accounting for 17% of the total. It is unsurprising that this is the second most common error type, as the category acts as a catch-all for several types of errors. [Omission of RANK: Clause-Phrase-Word-Morpheme CLASS: Noun, Verb, Adjective, Adverb, Preposition, Conjunction, etc] yielded 13 errors, accounting for 10% of all error types. Again, there was a great difference between the most and second most common categories. 11 errors (9%) were uncategorisable.

4.3.5.2 Ease of use

The headings in the rows and columns are clear enough and would not require further explanation, making this easy to use. Furthermore, it allows for more ease and depth of analysis, as now errors can be categorised as both level-type and modification-type. For example, *Wait *minute* is an error of [Omission] and a [Collocation] error. This avoids the problem of allocating the error to only one of these categories. There were very few uncategorisable errors (9% of all errors: one of the lowest numbers of all the frameworks). Most of these were [Paraphrase] errors that were neither awkward expression nor incoherent, and a category for these would have reduced the uncategorisable percentage to a negligible number

(2%). However, the issue of error count with [Paraphrase] errors remains: how should one count the number of errors in a sentence such as **all people there aren't in the globe next a few years*? I would argue that there would be three, based on the three phrases (noun *'all people there'*, verb *'aren't in the globe'* and prepositional phrase *'next a few years'*). Overall, there were relatively few instances of possible dual categorisation: errors with prepositions could have been categorised under [Class: preposition or collocation]. Other errors could also have been allocated to [Collocation] or the word class under which they were made. For example, *...population have some *goods <advantages*. This could be a [Collocational] or a [Noun word class] error. Because there are strong and weak collocations, which makes identification of collocation subjective, one option could be to remove the [Collocation] option and offer only word class categories. Removal of a [Collocation] category would also remove indecision over categorisation of phrases, such as *'I *highly consider that..'*, which could conceivably be considered a [Felicity] error. Felicity itself is quite subjective. Other count issues surround collocational phrases. Is **In the other side* one error or two? Spelling errors are dealt with systematically, but errors such as *... won't be enough food to *fed...* could be considered a [Wrong word] or grammatical error.

4.3.5.3 Depth of analysis

One advantage of James's system is that it allows for both quantitative error analysis and a record of an individual student's actual errors. There is space in the table to record the error and the line number (and also, perhaps, the corrected version). This has obvious uses for feedback to the individual learner. This principle of multi-dimensional view/categorisation of error could be expanded to include intra/interlingual analysis. Errors of [Under-specification] could not easily be allocated in this version of the framework. A more quantitative picture could be gained by separating the sub-sections into their own columns. Although there is a large number of categories, the analysis only reported errors in nine of them. What was gained in ease of use, seems to be lost somewhat in depth of analysis, which, of course, could be a feature of the sample used. A wider range of error types would have been reported had there been less bunching of types into one category. For example, the [Discourse misselection] category saw errors of [Incoherence], [Infelicity] and [Cohesion]. Ideally, these should be separated out to provide a more detailed picture of the type of errors being made.

4.3.5.4 Points to consider when creating a new framework for LEA

- Retain the dual axes approach to offer more depth of analysis.
- Separate out the rows to include more types of LE.
- Remove the [Collocation] option and offer only word class categories.
- Include a [Paraphrase] category.
- Count paraphrase errors by the number of erroneous phrases (i.e., noun phrase, verb phrase and prepositional phrase).

4.3.6 Hemchua and Schmitt (2006)

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
A Formal Errors							
1 Formal misselection							
A1.1 Suffix type	3	0	0	2	1	6	5
A1.2 Prefix type	0	0	0	0	0	0	0
A1.3 Vowel-based type	0	3	0	0	0	2	2
A1.4 Consonant-based type	1	0	0	0	0	1	1
A1.5 False friends	0	0	0	0	0	0	0
A2 Misformations	0	0	0	0	0	0	0
A2.1 Borrowings							
A2.2 Coinage	1	0	0	0	0	1	1
A2.3 Calque	3	4	3	3	6	19	15
A3 Distortions							
A3.1 Omissions	1	0	2	0	2	5	4
A3.2 Overinclusion	0	0	0	0	1	1	1
A3.3 Misselection	0	0	0	2	1	3	2
A3.4 Misordering	1	0	0	0	4	5	4
A3.5 Blending	0	0	0	0	0	0	0
B Semantic Errors							
B1 Confusion of sense relations							
B1.1 General term for specific one	0	0	0	0	0	0	0
B1.2 Overly specific term	0	0	0	0	0	0	0
B1.3 Inappropriate co-hyponyms	0	0	0	0	0	0	0
B1.4 Near synonyms	4	2	4	4	0	14	11
B2 Collocation errors	2	2	0	7	2	13	10
B2.1 Semantic word selection							

Category	P4	P5	P6	P12	P13	Total No Errors	% of all errors
B2.2 Statistically weighted preferences	0	0	0	0	0	0	0

B2.3 Arbitrary combinations and irreversible binomials	0	0	0	0	0	0	0
B2.4 Preposition partners	2	8	7	3	4	24	20
B3 Connotation errors	0	0	0	0	0	0	0
B4 Stylistic errors							
B4.1 Verbosity	1	4	2	0	2	9	7
B4.2 Underspecification	5	0	1	1	0	7	6
Total number of errors identified by the framework	24	23	19	22	23	111	89
Other errors							
Missing Word	2	1	2	0	0	5	4
Two words should be one	0	0	0	1	0	1	1
Completely wrong word	2	0	1	0	0	3	2
Unnecessary word	0	1	0	0	0	1	1
Paraphrase error	0	0	0	0	0	0	0
Miscellaneous errors with coherence of lexeme/ phrase.	2	0	0	1	0	3	2
Uncategorisable total	6	2	3	2	0	13	10
Total number of errors	30	25	22	24	23	124	100%

Table 25 Number of errors in each category and the number of uncategorisable errors using Hemchua and Schmitt (2006)

4.3.6.1 Discussion of results

To summarise, in this LEA, using this framework, a total of 111 error tokens were found across 24 error types. The most common LE here was the [Prepositional partner] error (24 accounting for 20%), the second most common was [Calque] (19 errors accounting for 15%) and then [Near Synonym] errors (14 errors, accounting for 11% of all errors). Again, given the large number of categories, there was a surprisingly large number of uncategorisable errors (13, accounting for 10% of all errors). One may have assumed that the larger the number of categories, the smaller number of uncategorisable errors. The large number of uncategorisable errors was largely due to a lack of a [Completely wrong word] or [Field] category, [Missing word] and [Miscellaneous errors with coherence of lexeme/ phrase] categories.

4.3.6.2 Ease of use

As described in Study 1, although many errors clearly belonged in a particular category, many could also have been categorised in different ways. This led to creation of a series of rules (e.g., [A2.3 Calque] error takes priority over other errors). These rules became numerous and reduced the potential ease of use of the framework for less linguistically aware teachers. Hemchua and Schmitt themselves report issues in their own study. Allocating errors to [A2.3 Calque] was

simple enough for the bilingual analyst, but more guidance is required on extent of error or error count. There was some indecision over allocation between [B1.4 Near Synonym] and [B2.1 Semantic word selection].

4.3.6.3 Depth of analysis

The specific naming of the categories and the relatively large number of them allowed for a relatively clear picture of the types of errors being made. However, mixing cause and type in one framework reduced the clarity of analysis, as doing so automatically created the possibility of dual categorisation, despite Hemchua and Schmitt suggesting that [Calque] takes priority. The highest number of errors were made with [Prepositional partners], but from the table above, it is not clear whether these were errors of [Omission], [Addition], [Selection]. There were several categories for spelling errors. Although this did not cause confusion, is not clear what advantage there may be to the learner, or SLA researcher in separating out spelling errors made with vowels and consonants or whether the error resulted in a word that exists in English or not. Interestingly, ten categories yielded no errors, e.g., [Connotation]. although this is unsurprising given the small sample size. 11% of all errors were uncategorisable, a quite high percentage, which was surprising given the number of categories. Most were [Completely wrong word] and [Missing word], suggesting that these two categories should be included in an improved framework for LEA.

4.3.6.4 Points to consider when creating a new framework for LEA

- Have only one category for [Spelling] errors.
- Combine [Near synonym] and [Semantic word selection] categories
- Include a [Missing word] category
- Include a [Completely wrong word] category
- Include an [Extra, unnecessary word] category
- Include an [Incoherence] category
- Include a [Paraphrase] category
- Include a [Two words should be one] category
- Remove cause categories
- Remove some categories, due to low frequency of errors

The following frameworks were not included in the current study for various reasons, but they do have some potentially useful features that may be worth incorporating into the design of a new framework.

4.3.7 Richards (1971) (Origin of influence or cause criterion)

Despite the fact that the framework was not used in the current study, for reasons given in Section 2.5.7.2, the following features, taken from Richards' study, are of interest when proposing a new framework as all of these features of Richards (1971) will be of value to teachers and students.

- Consider the inclusion of category for [Commonly confused words].
- Include categories for the [Omission], [Addition] of prepositions as well as errors made with [Misselection of individual, specific prepositions].
- Include a category for [Incorrect word ending for all words except verbs].
- Include categories for [Addition] and [Omission] of words/phrases.

4.3.8 Meara and English (1987)

Although the framework was omitted from the current study due to its similarity with other more comprehensive frameworks, the following points should be considered when designing a new, improved framework:

- Retain categories for [Totally wrong word] and [Wrong word, right semantic area] (as in Zimmerman 1986), as this would be an improvement on other frameworks that omit the former.
- Having a single category for [Spelling] makes analysis more straightforward, but subdividing into [Omission], [Addition], [Selection], [Ordering] would provide more depth.

4.3.9 Zughoul (1991)

Due to the high probability of dual categorisation discussed in Section 2.5.7.8 and the fact that the framework was devised with specifically Arabic learners in mind, it was not included in the current study. However, the framework shows some originality and there are some interesting points that could be used in a new framework:

- Categories 12 and 13, which describe [Infelicitous] and [Stylistic language] errors would also provide useful feedback to learners.

4.3.10 Lennon (1991)

This framework was also omitted from the study due to its primary focus on grammatical error analysis and part of speech, as opposed to the more frequent and perhaps more useful categorisation of types of LE, such as, [Coherence], [Verbosity] and [Overspecification] (i.e. informing learners that they are making errors with verbs is perhaps less useful in terms of constructive feedback than telling them that they are making errors with [Verbosity], for example: the latter gives them information they can act upon). However, one could consider some advantages of this framework when designing an improved framework for LEA: due to there being a finite number of parts of speech, there should be very few uncategorisable errors and a [Miscellaneous collocational errors] category would provide for this. Perhaps this explains Lennon’s claim that use of this framework results in very few uncategorisable errors. However, lexical chunks could not be classified this way.

4.3.11 Llach (2011)

Although this framework was not used in the current study, due to its similarity with other frameworks and a limited number of categories, there are some points that are advantageous and could inform the design of a new framework:

- Having only one category for [Misspelling] simplifies matters.
- Category 5, [Misselection], could encompass a variety of errors, including not only spelling, but also incorrect suffixes, which would imply a view of language as grammaticised lexis, as opposed to lexicalised grammar.

4.4 Comparison of results

Frame work	Category	P4	P5	P6	P 12	P 13	Total error tokens	% of all errors	Total error types
------------	----------	----	----	----	------	------	--------------------	-----------------	-------------------

Dušková (1969) (Etiologic/ process- oriented)	Errors identified by framework	19	9	5	14	6	53	42.8	4
	Uncat total	14	13	17	9	18	71	57.3	
	Total No errors	33	22	22	23	24	124	100	
Corder 1973 (Grammati cal or linguistic criterion)	Errors identified by framework	27	23	19	17	25	111	92	4
	Uncat total	3	0	1	5	1	10	8	
	Total No errors	30	23	20	22	26	121	100	
Zimmerma n (1986) (Descriptiv e/ product- oriented)	Errors identified by framework	33	19	18	24	22	116	91	11
	Uncat total	6	0	1	3	1	11	9	
	Total No errors	39	19	20	26	23	127	100	
Engber (1995) (Etiologic/ product- oriented)	Errors identified by framework	22	13	15	23	25	99	75	9
	Uncat total	13	6	4	5	6	31	26	
	Total No errors	35	19	19	28	31	132	100	
James (1998) (Form and content oriented)	Errors identified by framework	33	20	19	18	26	116	91	20
	Uncat total	2	1	2	5	1	11	9	
	Total No errors	35	21	21	23	27	127	100	
Hemchua and Schmitt (2006) (Form and content oriented)	Errors identified by framework	24	23	19	22	23	111	89	24
	Uncat total	6	2	3	2	0	13	10	
	Total No errors	30	25	22	24	23	124	100	

Table 26 Comparison of the number of errors identified by the various LEA frameworks and the numbers of categorisable vs uncategorisable errors

It can be seen from Table 27 that when using the different frameworks, different numbers of LE tokens and types were identified. This was due to the different ways in which the original authors classified LEs, or the categories used. For example, frameworks with categories for multiple errors in a phrase (e.g. Engber's with the highest yield showed 9 errors in this category.). Also, it must be said, I

performed the LEA once with each framework and was using each for the first time, so human errors due to inattention or fatigue could have been made. On reflection, I should have checked each analysis at least twice.

Despite the differences in the total number of errors identified by the different frameworks, these numbers are quite similar.

Zimmerman's (1986) and James' (1998) framework allowed for the identification of the most errors (116 each) and Dušková's (1969) for the fewest (53). This is quite surprising, given that Hemchua and Schmitt's (2006) had many more types (24). Dušková's (1969) yielded the most uncategoryable errors (57.3%) and Corder's (1973) the fewest, yet both had a similar number of categories, suggesting that it is not the number of categories that is important, but the naming of the type. James' (1998) framework yielded only 11 uncategoryable errors, yet had many more types, suggesting that his framework offered the best balance in terms of ease of use and depth of analysis.

4.5 Conclusions

The same five essays produced by Intermediate Greek learners of English were used to test a number of frameworks for their ease of use and depth of analysis, and the proportion of uncategoryable errors that they yielded. The advantages and disadvantages of the frameworks in these areas were considered.

There is some variation in the different frameworks in terms of the numbers of errors they reveal, the categorisation of errors, the ease of use, the depth of analysis that they provide and also in the number of uncategoryable errors that they produce. The frameworks have different approaches to analysis: some investigate cause, others seek to describe the errors only. A common method of categorisation in LEA frameworks is formal and semantic errors. It seems that whilst some are easier to use and provide more depth, whilst having fewer uncategoryable errors (James 1998), none are perfectly fit for purpose. Much can be learned from these analyses that will inform the design of a new framework with a more satisfactory balance between user-friendliness and depth of analysis that will yield fewer uncategoryable errors.

It seems that removing cause from the analysis reduces speculation over dual categorisation and, following James (1998), having two axes for categorisation (type of error and the surface taxonomy of [Omission], [Addition], [Selection], [Ordering] and [Blend] supplies a much more detailed picture than using only one

of these dimensions for LEA. Analysis by word class could result in fewer uncategorisable errors, but the value of categorising errors in this way may be of less value than categorising them by LE type (e.g., [Verbosity], [Prepositional partner], etc. It is also felt that combining spelling errors into one category would simplify matters without reducing depth of analysis.

Based on this information, I describe a new framework, LEA Framework 1 (NewLEAF1) in the next chapter, drawing mainly on the work of Leech (1981), James (1998) and Hemchua and Schmitt (2006). This framework will be thoroughly tested in subsequent work for ease of use, depth of analysis, and reliability.

Chapter Five (Study 3) Description and testing of a New Lexical Error Analysis Framework (NewLEAF1)

5.0 Introduction

The previous chapter examined six published frameworks for their ease of use, depth of analysis and useful aspects that could be incorporated into the design of an improved framework for analysing LEs. Following these analyses, it is clear that an improved framework needs, ideally, to be straightforward to use in terms of categorisation, which is not easily achieved, given the multi-dimensional, overlapping ways of viewing and categorising lexis itself. Whilst the literature shows several ways of analysing LEs, a system that incorporates the most useful aspects of these without causing confusion or dual categorisation must be found: a good framework would facilitate consistency and speed of recognition, as well as clarity, and even if the lexical stock of English is, by its nature, subject to constant adaptation and addition, making error identification problematic, there should be some guidance as to what constitutes an error. Perhaps more importantly, a new framework should offer a detailed picture, or as broad a spectrum as possible, of the type and frequency of LEs produced by learners. And finally, if consistency between analysts is to be achieved, guidance in the identification, counting and categorisation must be offered with brevity and clarity. The new framework appears to offer a balance between ease of use and depth of analysis.

This chapter will describe the new framework, explain the categories that have been selected, show how their selection is based on the advantages of previous frameworks used in the previous chapter, discuss the adaptability and limitations of the framework and offer guidance for identification, counting and categorisation of errors. Finally, the framework will be tested using the same essay data that was used in Chapters 3 and 4 (Studies 1 and 2), for comparative purposes. The results are compared with those obtained using other frameworks. In the next chapter (Chapter 6, Study 4), the framework will be tested for agreement of results when used by different analysts.

5.1 Description of NewLEAF1

With the requirements for LEA in mind, the proposed new framework (see Table 28 below) was designed after consideration of the advantages and disadvantages of previous frameworks for LEA. It draws on the work of Corder (1973), Leech (1981), Dulay, Burt and Krashen (1982), Zimmermann (1986), James (1998) and Hemchua and Schmitt (2006) in that it retains a large number of their categories for depth of analysis. The framework retains the distinction between formal and

semantic error analysis as this appears to be the most logical and least problematic approach, and with a large number of categories and two axes for analysis, should allow for greater depth of analysis than previously seen. It could be described in Llach's (2011) terms as form-oriented (Section A of the framework) and content-oriented (Section B). It could not be described as an origin of influence type, as it does not seek to organise errors by cause. Following Corder (1973), ascribing cause could be carried out at a later stage, as the two final columns allow for further categorisation of the LEs into interlingual and intra-lingual/development errors: it is not done here for reasons given in Section 2.4.5. The framework follows an expanded view of lexis (See Section 2.5.3), as it allows for the identification and categorisation of errors which may previously have been viewed as grammatical in nature, e.g., all prepositions, not simply prepositional partners and pronoun choice. Following James' (1998) recommendation, it has two axes for analysing error:

- a) A surface taxonomy. [Omission], [Over-inclusion], [Misselection] and [Misorder] following Corder (1973) and [Blend] following James (1998) have been put in the horizontal axis. This will allow for the subdivision of the types of error made with specific LEs in the vertical axis.
- b) A taxonomy of LE types (E.g., [Suffix], [Sense relations], etc.) has been put in the vertical axis. These are largely based on Hemchua and Schmitt's (2006) framework, which in turn is based on the work of Leech (1981), and Laufer (1991). It mainly retains their comprehensive categorisation of errors, which allows for an in-depth analysis.

However, compared with Hemchua and Schmitt's (2006) framework, there are now fewer categories in the vertical axis. For example, their A3.1-A3.5 categories have been combined to form a single category for [Spelling]. Although some spelling errors are more serious than others, it is not clear what value there was in Hemchua and Schmitt's framework for differentiating between spelling errors made with vowels and consonants, unless this was intended to help with spotting systematic spelling errors. One of the biggest differences between the new framework and previous ones is the absence of a category for [Collocation]. Omitting this avoids the difficulty in deciding whether a phrase is a collocation: therefore, the possibility of dual categorisation is also avoided. Previously one might have been undecided over whether to allocate an error to a collocation category or type of collocation category. Instead, collocation errors can now be

allocated to their more specific categories only, e.g., [Prepositions] or [Irreversible binominals], or placed in the miscellaneous, [Phrase] category.

This categorisation using two axes should offer more depth of analysis than using just one axis, and thereby offer a multi-dimensional view of error. This makes the framework easier to use by removing the possibility of allocation to more than one category on one axis alone.

The new framework does not rule out the possibility of interpretative discrepancies. To mitigate problems in this area, errors have been placed in my perceived rank order of seriousness within sections, with those most likely to impair communication nearer the top. In instances of possible dual categorisation, the analyst should choose the category which more specifically describes the error, or if there is further uncertainty, the error should be placed in the category nearer the top of the section. The assumption here is that communicative effectiveness is the basis for the evaluation of error 'seriousness'. This will also allow for more systematic and consistent inter-rater agreement.

In the current framework, there is also the option to record the number of error-free clauses in the writing, which when calculated as a percentage of the total number of clauses, will provide an accuracy figure. Some of the cells below contain 'NA' (not applicable). This is because it is not possible to have, for example, a misordering of a single suffix or prefix.

	OMISSION	OVER-INCLUSION	MIS SELECTION	MIS ORDER	BLEND	INTER-LINGUAL	INTRA-LINGUAL
Section A Formal Lexical Errors							
A1 Suffix Type				N/A			
A2 Prefix Type				N/A			
A3 Spelling							
A4 Two words should be one	N/A	N/A	N/A	N/A			
A5 One word should be two	N/A	N/A	N/A	N/A			
Section B Semantic Lexical Errors							
B1 Coherence (Words do not exist or phrase unclear)	N/A	N/A		N/A			
B2 Wrong Word (Field Error)	N/A	N/A		N/A	N/A		
B3 Wrong Word (Sense Relations Error)	N/A	N/A		N/A	N/A		
B4 Prepositions				N/A			
B5 Cohesion (Lexical substitution, pronoun use, conjunctions)							
B6 Paraphrase							
B7 Wrong Words (Statistically weighted preferences)							
B8 Arbitrary combinations and irreversible binomials	N/A	N/A	N/A				
B9 Connotative meaning	N/A	N/A		N/A	N/A		
B10 Verbosity/extra word/Repetition							
B11 Underspecification/missing word							
B12 Formality	N/A	N/A		N/A			
B13 Miscellaneous							

Number of error free clauses ___ out of ___ = ___% accuracy

Table 27 NewLEAF1

5.1.2 The categories

This section will describe the categories of the new framework below and offer examples where required.

5.1.2.1 Section A

[A1 Suffix type] and [A2 Prefix type] categories are retained from Hemchua and Schmitt's (2006) framework, as correct prefix and suffix use is important. Affixes often denote word class, and are relatively unproblematic to identify and allocate to category. They are near the top of the section, as affixes carry communicative value.

[A3 Spelling]. Following Corder (1973), Zimmermann (1986), Meara and English (1987), Engber (1995), James' (1998) and Llach (2011), there is now only one category for spelling error. This is to reduce the number of categories. Further, the benefit of separating errors with consonants and vowels is unclear (as in Hemchua and Schmitt 2006). To allocate to this category, the meaning of the word that contains the error is clear and those words exist in English. Words whose meaning is not clear should be categorised under [B1 Coherence].

[A4 Two words should be one] and [A5 One word should be two]. The evolution of morphology has seen two words become hyphenated and then become a compound noun, e.g., wastepaper. Identification of where certain lexical items may be on the scale of how fixed they are as one or two words: e.g., clothes brush, toothbrush and horse-brush) and may cause disagreement between analysts. There are instances when there are obvious, widely agreed errors in these areas (e.g., **tooth brush*). Only these should be included in the analysis.

5.1.2.1 Section B

[B1 Coherence] (words do not exist in English or meaning of phrase or words is unclear). It is sometimes the case that the learner's writing contains so many or such serious errors that meaning is unclear. If one cannot decipher what the intended meaning is, then it becomes impossible to categorise specific erroneous words or clauses. Attempts at doing so would produce inaccurate results or

skewed data. Omission of a [Coherence] category would lead to the exclusion of serious errors in the analysis. Inclusion of this category allows for the inclusion of whole phrases as incoherent and enables error analysis to be performed on largely incomprehensible writing. A counter-argument here is that even though a succession of errors categorised as 'incoherent' are not in themselves of much analytical or pedagogical value, categorisation will help to isolate them for attention. An incoherence category was oddly missing from several of the frameworks that were examined. This is surely one of the more important categories to include, given the view that error seriousness should be based on communication breakdown, hence its position at the top of the section.

[B2 Wrong word (field error)] This category was also oddly absent in Hemchua and Schmitt's (2006) and other frameworks. However, Zimmerman (1986), Meara and English (1987) and Engber (1995) included a [Totally wrong word] or [Field error] category. This category is for when the error word does exist in the L2, but cannot be used in this sense, they are not of the same lexical set or meaning group, or are not involved in the sense relations of the word that should have been used. For example, the error could be a 'False-friend' (e.g., *I feel *serious* <stressed>). Compare with [B3 Wrong word (sense relations error)], which is for related words.

[B3 Wrong word (sense relations error)]. Following Dušková (1969), Zimmerman (1986), Meara and English (1987), Engber (1995), James (1998) and Llach (2011), these various sense relations categories were combined into one. This reduces the number of categories and therefore the possibility of dual categorisation, and also makes the framework easier to use by less linguistically aware teachers who may not be comfortable discriminating between concepts such as antonymy, synonymy, superonymy, (co-)hyponymy and meronymy. This category is for lexis that exists in English and is in the same lexical set as the intended target word.

[B4 Preposition]. Preposition partner errors seems to be one of the most common types of errors (See Studies 1 and 2). The category was retained from Hemchua and Schmitt (2006), but expanded to include all preposition errors, such as **In the other hand*, the preposition error as part of a collocational phrase, and errors with prepositions of time and place, e.g., *See you *in the weekend*. Although

prepositions of time and place have traditionally been taught under the heading of grammar, a more modern view of lexis could see them as more lexical in nature (Lewis 1992). Furthermore, they are included because it is easy and useful to do so. Categorisation is unproblematic.

[B5 Cohesion] (lexical substitution, pronoun use, conjunctions). James (1998) classified these as discourse errors. While this may be accurate, errors in lexical substitution could also be identified at the word or phrase level, (e.g., Dr Smith/he/the doctor are all lexis that could be used interchangeably to describe the same person or concept). It would also be beneficial to include this to provide a more comprehensive picture of types of learner error. Many would argue that pronoun use is in the realm of grammar, but they could also be seen as correct or incorrect lexical choices, given that they are words that carry specific meaning. The same could be said for conjunctions. They are individual lexemes that have their own meaning.

[B6 Paraphrase]. Following Zimmerman (1986 and 1987), Lennon (1991) and Engber (1995) there is now a category for phrase errors: whole problematic phrases which require re-writing, e.g., *next a few years. They could simply be instances of awkward expression. Lennon (1991) discussed the distinction between domain and extent in error analysis to overcome issues where whole parts of sentences were understandable in context, but erroneous for more than one reason. It is a catch-all category to be used when errors do not fit into categories below. The meaning is clear. These phrases could be mainly erroneous due to [Omission], [Over-inclusion], [Misselection] and [Misorder] of more than one lexical item. The inclusion of this category can be seen as problematic, but it is a way of dealing with error identification in some instances, and if we break all errors down to individual words, we are ignoring the formulaic nature of language. As we have seen, erroneous language does not exist solely in single words, but can manifest across words within a clause, making the whole clause difficult to comprehend or seem incorrect. Identifying two or three of these words may be insufficient to correct the error to a satisfactory degree and the correction may not reflect the learner's meaning entirely. For example, the clause, *all the people there (context tells us that the author meant 'the global population') cannot be corrected by changing some of the words. The correct form is a collocation and a

single meaning unit. It would be time inefficient to try to unpick which words are incorrect, missing or in the wrong order and preferable to teach/explain the correct form as an unanalysed whole or chunk of language. Although a focus on phrase-level error may appear to lack the specificity of single-word feedback, and therefore make it harder for students to act on feedback, it does reflect the phrase-based nature of language (Wray 2008).

[B7 Statistically weighted preference], [B8 Arbitrary combinations and irreversible binomials] and [B9 Connotative meaning]. Following use of Hemchua and Schmitt's (2006) framework, these were found to be rare, but the categories were retained as allocations to categories here would be clear and unproblematic and provide for a greater range of error types, should they occur.

[B10 Verbosity/ Unnecessary words/ Repetition] and [B11 Underspecification /Missing words.] These two sets of areas were grouped together, as at times it is difficult to differentiate between them. These sections were included for phrases that are lexico-grammatically correct, but require fewer words or more detail to clarify meaning, or simply to improve expression. E.g., *He bought an apple and* he bought a banana.* <He bought an apple and a banana. This parataxis is rare outside the deliberate stylistic effects associated with creative writing (e.g., the works of Ernest Hemingway or Cormac McCarthy).

[B12 Formality]. For words or phrases that are overly formal or informal for the genre in use. This category was included to allocate lexico-grammatically correct phrases, with the wrong level of formality. Since sensitivity to register in the foreign language develops over time, errors with formality can be frequent. Oddly absent from a number of LEA frameworks as it is more of a sociolinguistic phenomenon. E.g., **I informed my girlfriend via the medium of the telephone.* <I told my girlfriend on the phone.

[B13 Miscellaneous]. This was included for any errors that do not fit into the categories above. It is recognised that other types of errors not encountered with the current data set may occur, and that despite refinements to the framework, some errors may defy categorisation. This category was included to ensure all errors are counted so that even if the error cannot be categorised, it is included in

the analysis, and also to investigate the miscellaneous errors so that the framework can be evaluated and further refined by adding appropriate categories in the future.

At the bottom of the LEA framework is a space to record the number of error free clauses. This information could be used to provide an accuracy ratio when presented with the total number of clauses. One problem associated with EA in general is that it does not give credit for what the learner can do well. A good ratio would mitigate this issue. Further, it has not been claimed here that LEA is a 'one size fits all' solution for evaluation and feedback. Instead, it should be seen as an available tool to be used when a focus on accuracy is required. Of course, specific praise could also be provided qualitatively. A linguistically sensitive teacher could, for instance, praise the avoidance of false friends, rich vocabulary resource, variation of sentence lengths, attention to helping the reader via signposting strategies or any number of other areas of writing.

5.2 Variation - whole language analysis, including grammar, phonology and punctuation

While the framework primarily focuses on lexis, this does not necessarily mean that other areas of language need to be excluded. The framework is adaptable. With the current horizontal axes of surface error taxonomy, extra categories for analysing grammar, pronunciation and punctuation could easily be added into the framework by including more rows detailing the construct to be analysed.

5.3 Adaptability

There is a degree of adaptability. For instance, if complete accuracy is not a goal of language instruction, and the teacher wishes to foreground communicative ability and avoid possible demotivation via a high error count, some categories, which seem to demand a high level of accuracy, could be omitted from the framework. For example, [B7 Statistically weighted preference], [B8 Arbitrary combinations, irreversible binomials] and [B9 Connotative Meaning]. In addition, when the tutor/researcher is conducting analyses on multi-L1 groups' work, or he/she cannot speak the students' L1, the final two columns, [Inter-lingual] and [Intra-lingual] could be deleted, thereby making this a description-only taxonomy of errors.

5.4 Analysis

To establish whether the framework is easy to use and whether it provides depth of analysis, for comparative purposes, I tested it on the same five compositions that were used in Study 2, using the process guidance outlined below.

5.4.1 Guidance for analysts using the new LEA framework

This section discusses guidance for analysts using the framework in the areas of identification, counting and categorisation of LEs. Some specific advice is required to help the analyst understand how to use the framework and ensure inter-rater agreement. The specific guidance accompanying NewLEAF1 can be found in Appendix 5.1. During the analyses and completion of previous chapters, working with previously published frameworks proved to be a little frustrating due to the lack of accompanying guidance in terms of:

- a) **error identification** What constitutes an error and a LE? Llach (2011) also bemoans this lack of clarity in some previous studies.
- b) **error counting** Should one attempt to differentiate between a mistake and error? Should repeated errors be counted once or twice? Should multiple errors in a collocational phrase be counted as one or multiple errors?
- c) **error categorisation** Clarity here relies on clear examples and explanations of the type of errors that can be categorised using the framework. There should also be guidance as to what to do in the case of the possibility of dual categorisation.

These issues led to deliberation of whether an error had been made, whether it was grammatical or lexical in nature, how many had been made and how to categorise them. Had all of the various authors offered more guidance in these areas, then results from the various analyses may have been more comparable. To overcome issues in these areas, the current framework will be published with the following advice that attempts to clarify these issues. It is hoped that this advice will be of particular use to less-experienced teachers or those with less linguistic training.

5.5 Results

This section reports the results of the LEA when using NewLEAF1 on the same five compositions that were used in Study 2, and compares these results with those from previously published frameworks (results extracted from Study 2).

5.5.1 Results of the lexical error analysis

Overall, as can be seen in Table 29 below, semantic errors far outnumbered formal ones (88% and 12% respectively). The most common type of error identified in the five essays was [B4 Wrong words (field error)]. There were 25 errors in this area, accounting for 20% of all errors. The joint second most frequent error types were [B7 Prepositions] and [B3 Wrong words (sense relations)], both yielding 20 errors, each making up 16% of all error types. These error types were unevenly distributed across the five essays: as predicted, there were no or very few errors made in several of the categories. Errors appeared in approximately half of the categories (15 out of 34 categories). Of these 15 categories, all students made errors in five of them, four students made errors in two of them, three students made errors in one of them, two students made errors in three of them and one student made errors in five of them. This suggests that these students make similar errors to each other.

Category	P4	P5	P6	P12	P13	Total errors	% of all errors
Section A Formal Errors							
A1 Suffix type							
Omission	0	1	0	0	0	1	0.8
Over-inclusion	0	0	0	0	0	0	0
Misselection	3	0	0	1	0	4	3.2
Misorder	0	0	0	0	0	0	0
Blend	0	0	0	0	0	0	0
A2 Prefix type							
Omission	0	0	0	0	0	0	0
Over-inclusion	0	0	0	0	0	0	0
Misselection	0	0	0	0	0	0	0
Misorder	0	0	0	0	0	0	0
Blend	0	0	0	0	0	0	0
A3 Spelling							
Omission	2	0	1	1	2	6	4.8
Over-inclusion	0	0	0	0	0	0	0
Misselection	0	0	0	0	0	0	0
Misorder	0	0	0	0	3	3	2.4
Blend	0	0	0	0	0	0	0

A4 Two words should be one - Blend	0	0	0	2	0	2	1.6
A5 One word should be two	0	0	0	0	0	0	0
Totals						15	12
Section B Semantic Errors							
B1 Coherence (Words do not exist in English)	5	0	0	0	0	5	4
B2 Wrong Words (Field)	7	5	2	5	6	25	20
B3 Wrong Words (Sense Relations)	4	1	4	8	3	20	16
B4 Prepositions							
Omission	0	0	0	1	1	2	1.6
Over-inclusion	0	0	0	0	1	1	0.8
Misselection	1	6	7	2	1	17	13.6
Misorder	0	0	0	0	0	0	0
Blend	0	0	0	0	0	0	0
B5 Cohesion (Lexical substitution, pronoun use, conjunctions)	0	0	0	1	1	2	1.6
B6 Paraphrase	6	2	4	3	3	18	14.4
B7 Wrong Words (Statistically weighted preferences)	0	0	0	0	0	0	0
B8 Arbitrary combinations and irreversible binomials	0	0	0	0	0	0	0
B9 Connotative meaning	0	0	0	0	0	0	0
B10 Verbosity/extra word/Repetition	5	4	2	1	1	13	10.4
B11 Underspecification/missing word	0	2	3	1	0	6	4.8
B12 Formality	0	0	0	0	0	0	0
B1 Miscellaneous	0	0	0	0	0	0	0
Totals						109	88
Other errors	0	0	0	0	0	0	0
Uncategorisable total	0	0	0	0	0	0	0
Total number of errors	33	21	23	26	22	125	100%

Table 28 Results of LEA on Greek Data using NewLEAF1

To summarise, using this framework, 125 error tokens were found across 15 error types, the highest number of tokens and the third highest number of types when compared with the other frameworks in Table 27. The third highest number of types can be considered satisfactory when one considers that NewLEAF does not include causality categories. There was almost no deliberation over dual-categorisation and there were no uncategorisable errors found. The framework offers a wide spectrum of LE types and does not employ too much specific linguistic terminology, making it more user-friendly for a wider range of users. For example, there is no mention of super-ordinate, co-hyponyms, etc. However, it is predictable that when working with only one relatively small data set, one could

create a framework that yields no uncategorisable errors by tailoring the framework to that data set. However, the lack of such uncategorisable errors was mainly achieved by the provision of the [Paraphrase], [Coherence] and [Miscellaneous] categories. Eliminating the cause of error and [Collocation] categories helped to avoid dual categorisation.

5.5.2 Comparison of performance with other frameworks

For comparative purposes, the results of the above analysis are compared with those from the previous chapter (see Table 30 below) when using the same essay data. The number of categorisable and uncategorisable errors are presented. It can be seen that the new framework identified a similar number of errors in comparison to the results from other frameworks (124). However, all of the errors found with NewLEAF1 were categorisable.

Framework	Category	P4	P5	P6	P12	P13	Total Errors	% of all errors
Dušková (1969)	Total errors	19	9	5	14	6	53	42.8
	Uncat total	14	13	17	9	18	71	57.3
	Total	33	22	22	23	24	124	100
Corder (1973)	Total errors	27	23	19	17	25	111	92
	Uncat total	3	0	1	5	1	10	8
	Total	30	23	20	22	26	121	100
Zimmerman (1986)	Total errors	33	19	18	24	22	116	91
	Uncat total	6	0	1	3	1	11	9
	Total	39	19	20	26	23	127	100
Engber (1995)	Total errors	22	13	15	23	25	99	75
	Uncat total	13	6	4	5	6	31	26
	Total	35	19	19	28	31	132	100
James (1998)	Total errors	33	20	19	18	26	116	91
	Uncat total	2	1	2	5	1	11	9
	Total	35	21	21	23	27	127	100
Hemchua and Schmitt (2006)	Total errors	24	23	19	22	23	111	89
	Uncat total	6	2	3	2	0	13	11
	Total	30	25	22	24	23	124	100
NewLEAF1	Total errors	33	21	23	26	22	125	100
	Uncat total	0	0	0	0	0	0	0
	Total	33	21	23	26	22	125	100

Table 29 Comparison of the number of categorisable vs uncategorisable errors identified by Version 1 and previously published frameworks

5.6 Summary/conclusions

The current chapter has described, and provided a justification for, the design of a new framework. The design of this framework is based on the previous studies and the analysis of previous frameworks. Accompanying, detailed recommendations in terms of guidance for the analysts have been provided in the areas of identification, counting and categorisation of LEs. For comparative purposes, the new framework was tested using the same data from Greek learners that were used in Studies 1 and 2.

5.7 Limitations of the new framework

Finally, it must be mentioned that no system for LEA will be perfect, given some of the previous issues with error analysis in general, and that the most that can really be achieved is minimising the weaknesses of previous attempts to make the process as accurate as possible. Also, a purely quantitative approach is subject to natural limitations. Despite the suggestions offered in the section on variation, it is still felt that some qualitative comment is required to address other areas of writing, such as organisation or structure. The proposed framework is not designed to replace such qualitative feedback, however. Nor has LEA been proposed as a 'one-stop shop' for feedback on writing. However, the new framework goes a long way to addressing previous issues. In terms of error count difficulties: Dušková (1969:14) states that 'Fortunately, the number of cases in which it was hard to decide whether or not an error had been made...did not exceed 4% of all the errors made'. This seems to be an acceptably small figure: one which was not exceeded in the current research. In terms of categorisation, Lott (1985:259) states there is difficulty 'in building a system of definitions where the analyst will not periodically have doubts about how to categorise particular errors: the quality of any research must be affected by the researcher's intuitions about [the type of error and] why the error occurred'. This this issue has been greatly minimised with the new framework and a negligible number of instances where dual categorisation could occur is also to be expected, and such a small number could also be seen as acceptable.

It was found that NewLEAF1 uncovered a similar number of errors and no uncategorisable errors in comparison with previously published frameworks, which can be seen as good, but the former finding is also unsurprising, given that it was

designed primarily using that same data set. There were also no instances of possible dual-categorisation, showing that at least I found my own system usable. As was to be expected, with a different set of categories, the findings of the analysis was different from previous analyses. This limits the usefulness of comparison of findings.

Whilst it is positive that the framework yielded a similar number of errors, a broader spectrum of error types and no uncategorisable errors in comparison to previous frameworks when used by the designer, it does not guarantee that problems will not be found when analysis is performed by others. Any shortcomings of the new framework in terms of inter-rater agreement will be reported in the next chapter (Study 4) where it will be thoroughly tested by six highly-experienced and highly-qualified university EAP teachers.

Chapter 6 (Study 4) Testing and refining NewLEAF1

6.1 Introduction

Study 1 was a full replication of Hemchua and Schmitt (2006): the study showed similar findings to the original, which demonstrated that learners of different L1 groups may indeed make similar kinds of LEs. Study 2 tested six different published LE frameworks with the same data used in the replication study to ascertain their ease of use and depth of analysis provided. The disadvantages and advantages of the different frameworks from these two investigations were noted and used to devise NewLEAF1, which was described in Study 3 and tested on the same Greek data that were used in Study 1. It was found to be an improvement on previous frameworks. The current chapter details Study 4. It seeks to trial the new framework, gauge its inter-rater agreement and gain qualitative feedback on its potential for use, ease of use and depth of analysis that it provides. Results will be used to further refine the framework and produce a second version, which will be tested again in Study 5 to improve it further and produce a further refined version.

This chapter starts by introducing the specific research questions to be answered and then describes the participants, ethics and methods of data collection and analysis. Results are presented and recommendations for improvements to the framework are made, based on these results. The refined framework (NewLEAF2) can be found in Appendix 6.1.

6.1.1 Aims

The study reported in this chapter aims to trial the new framework with six highly-experienced, highly-qualified English Language teachers who are also highly-skilled users of English. The study also aims to ascertain the inter-rater agreement of the framework, any issues they experienced whilst using the framework and the degree to which these teachers are satisfied with its depth of results and ease of use, and also how the results of LEA may be used in their line of work. This information is sought so that the framework and accompanying guidance can be improved. Therefore, the research questions are as follows:

6.1.2 Research questions

- 1) Is there an adequate level of inter-rater agreement when experienced language teachers are asked to identify LEs in a piece of student writing and categorise extracted LEs using the new framework?
- 2) What thoughts and opinions did participants have on the ease of use of the framework and what problems did they encounter when using the framework?
- 3) How satisfied are the teachers with the depth of analysis?
- 4) How might they use LEA results in their practice?
- 5) Based on this information, how might the framework and guidance be improved?

6.2 Methodology

Research Question	Data set used to answer questions	Qual / Quant	Purpose
1) Is there an adequate level of inter-rater agreement, when experienced language teachers are asked to identify LEs in a piece of student writing and categorise extracted LEs using the new framework?	Annotated scripts Completed LEA frameworks Screen capture	Quant	Trialling Establish inter-rater agreement
2) What thoughts and opinions did participants have on the agreement and ease of use of the framework and what problems did they encounter when using the framework?	Screen capture Semi-structured interviews	Qual	Establish ease of use
3) How satisfied are the teachers with the depth of analysis?	Semi-structured interviews	Qual	Establish depth of analysis
4) How might they use LEA results in their practice?	Semi-structured interviews	Qual	Establish application
5) Based on this information, how might the framework and guidance be improved?	Annotated scripts Completed charts Screen capture Semi-structured interviews	Qual	Trialling

Table 30 Research questions, data sets to answer questions and purpose

This section will explain the ontology of the current study, ethical considerations, the participants involved and how the data were collected and analysed. Table 31 above summarises the various data sets used to answer research questions.

6.2.1 Methodological Approach

Following my own philosophical research position, neither a quantitative only, nor qualitative only approach would paint a clear picture of the value or efficacy of the new framework. Neither would provide complete truths in this area (Cohen *et al.*, 2018). Furthermore, the research questions themselves could not be answered by one type of data collection alone. Therefore, a mixed methods approach was decided upon. This would combine the two to provide richer data sets, and therefore provide more detail, understanding and improve the findings' validity (Creswell and Plano, 2011 and Cohen *et al.*, 2018). Following Dörnyei (2007), qualitative design was incorporated to record and understand the teachers' responses to questions and opinions on the framework, which would help to unveil some of the complexities that may not have occurred to me during the design stage (Burton and Bartlett, 2009). I also felt that the recording of qualitative data, i.e., practising teachers' thoughts on the efficacy of the framework could be used to counter my own bias in this area (Denscombe 2014). It was also important to integrate a quantitative approach to compare the results of the analysis between participant teachers, as similarity of numerical results would have confirmed that the framework has good inter-rater agreement.

6.2.2 Ethical considerations

Published guidelines (2016) for conducting research at Manchester Metropolitan University were followed. Full ethical approval (approval number 0906) was gained from the University Ethics Committee (see Appendix 6.2 for letter of approval). The rights of participants (privacy and protection of personal data) were protected by provision of a Participant Information Sheet (see Appendix 6.3). This, in line with the General Data Protection Act 2018, explained how and why necessary personal data would be collected, stored and used. It also provided information about the study itself. Full anonymity was assured as participant numbers were used instead of participant names. A consent form was provided. This allowed all participants to approve the processing of their personal data (see Appendix 6.4 and 6.5 for student and teacher consent forms).

6.2.3 Participants

This section will offer a rationale for how participants were selected for this study and then describe the groups of participants, the students from whom scripts were collected and the teachers who analysed those scripts using NewLEAF1.

6.2.3.1 Sampling

The participants were found through convenience sampling. This involves utilising members of the research population that are the most easily accessible to the researcher (Given, 2008). This convenience sampling approach is not always ideal as it can compromise credibility: a positive working relationship between researcher and some of the participant teachers (described below) may tempt the latter to respond favourably towards the framework. This potential disadvantage was mitigated by asking them to point out issues, as it would be more helpful in the design of the framework. Furthermore, I was aware of the professionalism of these participant teachers, and their willingness and engagement were felt to be positive contributions to the research process and would result in a rich data set (Dörnyei, 2007). Six teachers were invited to participate, as a well-designed qualitative study can uncover rich data from a small number of participants (Dörnyei, 2007).

6.2.3.2 Students

All thirty-two students on a university pre-session course were invited to submit their essays for LEA feedback. They were aged between 20 and 28, both males and females. At the beginning of the course, they were proficient to IELTS 5.0 and were of mixed L1 groups (Chinese, Saudi, Kuwaiti, Indonesian and Ivorian).

6.2.3.3 Teachers

Experienced English Language teachers from a university in the Northwest of England were invited to participate in the research. Six responded. They were aged between 25 and 60 years old. The participants had between 10 and 30 years of teaching experience, post CELTA, which meant they were suitably experienced in responding to learners' written work, had experience of dealing with learners' LEs and would therefore be in a position to comment usefully on various aspects of the framework, guidance, its implementation and potential use for their practice. They were all L1 speakers of English, but this was not a pre-requisite. Instead, highly skilled users of English were sought.

All held either a master's degree in TESOL or Applied Linguistics and a Diploma in English Language Teaching to Adults (DELTA), all of which typically have a strong language awareness element. Therefore, they all had a suitable level of training in linguistics. These qualifications suggest that these participants should be familiar with the terminology used in the framework and guidance.

6.2.4 Data collection

This section will describe how the learners' scripts were collected and also how the participant teachers conducted the LEA of those scripts. This would provide the basis for LEA observation and interview to provide data to answer the five research questions above.

6.2.4.1 Collection of the scripts

The students, as part of their pre-session course, were asked to write approximately 350 words on the following brief: 'Reading is the most important academic skill. To what extent do you agree or disagree with this?' They had 40 minutes to complete the task, and did not have recourse to the internet or a dictionary. Essays were handwritten to avoid use of the spell and grammar checker in MS Word. Students were then invited to participate in the research by giving their permission for me to request the essay from the Pre-session Course Leader. The incentive was that they would receive detailed feedback on the LEs that they made. Five students responded. Their essays were carefully typed into Microsoft Word with the spelling and grammar checker turned off. One essay was selected for the current study (see Appendix 6.6). The criteria for selection were that there was a larger number of LEs and a wide range of them, compared to other essays, to satisfactorily test out the new LEA framework.

6.2.4.2 Lexical error analysis of the script

Table 32 provides a summary of the stages of data analysis. These are discussed further in subsequent sections. Participants were emailed the guidance, framework and accompanying instructions (Appendix 6.7) before the data collection took place. They were not given any additional verbal instructions as to how to perform the LEA to replicate real-life usage conditions: if consistency in LEA is to be achieved in different contexts, analysts will have to rely on the same written instructions without further guidance from myself. It was decided to focus on error

identification and categorisation, as these aspects, according to Ellis, (1994), are the two most problematic stages of error analysis.

1. Participant analysts were emailed the guidance, framework and accompanying instructions with no additional verbal instructions.
2. Paired participants worked together to identify all LEs highlighting the digitised scripts. I recorded their discussion in a 'paired think aloud' procedure. Kaltura desktop capture was used to video the screen and record their voices during the discussion. Pair work was also observed.
3. Paired participants categorised 37 LEs extracted from the essay and allocated error codes to these from the framework, also using a 'paired think aloud' procedure. Discussion was recorded with Kaltura. Pair work was observed.
4. Participants allocated a confidence score of 1-3 to their categorisations to show how certain they were about them.
5. My own LEA of the script was added to the data set to give four annotated scripts and seven sets of categorisations and accompanying confidence scores.
6. Participants were interviewed individually.

Table 31 Summary of stages of data analysis

Task 1

The six participant teachers were given written instructions to complete two tasks (see Appendix 6.7) and put into three pairs, given the digitised essay in Appendix 6.6, and then asked to identify all LEs. The three pairs of participants were asked to work together to identify all LEs by highlighting the digitised scripts. This was done so that they could be recorded discussing their completion of tasks for later analysis. Llach (2011) states that any LEA procedures should start with a definition of what constitutes a LE. Therefore, the definition in Section 2.5.5, based on George (1972) and Lennon (1991), was used. To assist participants in unpicking lexical from grammatical errors, which, according to Hemchua and Schmitt (2006), can be problematic, further advice was adapted from Hemchua and Schmitt (2006). (See Appendix 5.1.)

Task 2

Participants were then asked to discuss in pairs, but individually categorise the 37 LEs extracted from that essay by allocating error codes, taken from the framework

chart (see Table 1 in Appendix 6.7). Participants were also asked to individually allocate a confidence score of 1-3 to their categorisations to show how certain they were about them. Confidence scores would allow me to understand how sure participants were about the framework overall and their confidence in allocating individual errors to categories. A confidence score of 3 meant that participants felt sure that they had allocated the error to the only possible category. A score of 2 meant that they felt the error could have been allocated to more than one category. A score of 1 meant that they felt completely unsure about their categorisation.

It was decided to separate LEA data collection into these two tasks because if participants had been asked to identify LEs and then categorise those same LEs, they may have identified different ones from each other, making comparison of their categorisations very difficult.

My analysis was added to the data set to give four annotated scripts and seven sets of categorisations and accompanying confidence scores.

6.2.4.3 Think-aloud protocol

Think aloud protocol was used as I wanted to understand the thinking behind their decision making and iron out any confusion that transpired.

Participants were asked to think aloud in pairs to discuss and justify their categorisations and confidence scores. Kaltura desktop capture was used to video the screen and record their voices during the LEAs. If, whilst completing a task, the subjects fall silent for more than three seconds for any reason, it is possible to ask the subjects to say what they are thinking. However, this may interrupt the thinking of the subject and such prompts should be used sparingly. Therefore, to avoid some of the issues associated with think aloud protocol, the participants were asked to complete the analyses in pairs. This would have the following advantages: it would allow participants to schedule and complete the task in their own time and they would not have to wait for me to be free at that time also. However, it was decided that I should be present to ensure the data collection went smoothly. Further, as participants would have to agree before categorising errors, there is little risk of them falling silent. It is also a more efficient method as it collects data from two participants at once and provides the researcher with ready prepared discussion for analysis. One disadvantage of this approach is that one

analyst may dissuade the other from their preferred choices. However, this could also be seen as an advantage as the discussion would also help to clarify their understanding of the framework and guidance when performing LEA.

Think aloud protocols ask subjects to verbalise their thought processes whilst completing some kind of task and can expose detailed task-specific process models. According to Azevedo *et al.* (2015:764), think aloud protocol or 'protocol analysis' is a rigorous methodology for eliciting verbal reports of thought sequences of data on thinking, reasoning, problem solving, reading, and learning.' They also state that think aloud data, after it has been coded, can lead to 'experimental manipulations', or improvements to a process. Azevedo *et al.* (2015:764) also state that the think-aloud protocol, although time-consuming and requiring precision, has been popular for over 60 years since the cognitive revolution and information processing theories of psychological phenomena emerged in the 50s', and it has led to revolutionary styles of verbal reporting, much improved on other introspective methods, such as retrospective protocols. It has now become the dominant method of collecting data in protocol analysis, as it is often used to record and analyse the 'underlying cognitive processes and structures' that differ between subjects.

"This focus on simultaneous thinking aloud is a strict requirement of protocol analysis in order to identify the contents of [a subject's] working memory as directly as possible. The contents of working memory are of primary interest because they are associated with executive functioning. Working memory is the memory location where attentional resources are allocated when individuals seek to solve problems and, therefore, this is the memory location of interest for measuring (and acquiring evidence to support inferences) about a [subject's] problem- solving processes."
(Ericsson and Simon, 1993, in Azevedo *et al.*, 2015:764)

However, concurrent think-aloud protocol data collection is not without its issues. For example, Azevedo *et al.* ask whether explaining one's thought processes changes the way one approaches a task. This is known as 'reactivity': i.e., 'if think-aloud protocols change the cognition of interest, then verbal reports may misrepresent one's understanding of the contents of individuals' minds' (2015:764). However, Ericsson and Simon (1993 in Azevedo *et al.* 2015:764) state that verbal protocols do not change people's cognitive processes and are

relatively complete accounts of cognitive processes. Azevedo *et al.* (2015:764) also ask whether this particular type of elicitation produces different types of data from other methods. Nonveridicality may also be an issue in that asking people to verbalise their thoughts may not compel them to state the cognition of interest to the researcher, i.e., asking subjects to report how they believe they solved a task is their belief, and may not expose their task-solving processes in working memory. However, Ericsson and Simon (1993 in Azevedo *et al.* 2015:764), state that if concurrent think alouds are conducted as they prescribe, these issues can be avoided. A further issue with concurrent think-aloud data collection is the analysis of the data. This will be discussed in Section 6.2.5.

6.2.4.4 Observation of participants

On its own, observation would not give access to the thought processes in action when completing the tasks. Azevedo *et al.* recommend triangulating the think-aloud data with other process data, such as observation of participants, while they complete the LEA. This would 'add additional context that will facilitate inferences about the underlying cognitive phenomena' (2015:764). For example, silent pauses in the analyses may be explained through observation. Therefore, the participants were observed and monitored for body language and facial expression, which may indicate lack of comfort or otherwise with the procedure. These para-linguistic features were noted, along with the time (synchronised with the Kaltura screencast) and were used to add more context to comments made during the think aloud data.

6.2.4.5 Semi-structured individual interviews

Post analysis interviews were used, to answer subsequent research questions. They were included to triangulate data from other methods, as on their own, they would have been considered unreliable in that participants may not have been able to recall why they had made certain decisions during the analysis. Interview questions are often carefully crafted to focus on a specific topic or scenario and are often sequenced in a meaningful order, but may not touch upon what other areas the subject may have been thinking about. The interview schedule in Appendix 6.8 was used to gain deeper understanding of the participants' thoughts on the framework. Individual semi-structured interviews were conducted with three out of the six individual participants (one from each pair) to discuss areas relating to research questions 2, 3, 4 and 5. Of the various possible methods of

triangulating the data collected from the think aloud protocol, semi-structured interviews were chosen ahead of the following: structured, unstructured and focus group meetings. Structured interviews would be too rigid in nature and it would be difficult to adapt the schedule of questions to the issues that the individual participants experienced during the analyses. Unstructured interviews were rejected on the grounds that specific details about the ease of use of the framework were required and focus group interviews were not considered due to the difficulty of gathering all of the participants together at one time. Also, original individual opinions were sought to ensure that all participants' voices were heard and avoid potential dominance by one speaker in a focus group (Dörnyei, 2007 and Cohen *et al*, 2018).

6.2.5 Data analysis

This data analysis section starts with a summary of how the data were analysed. It is then organised by research question. It will describe and justify how the data were analysed and will also address associated issues with this type of data analysis. To summarise:

- 1) The identifications of LEs and the categorisations of the extracted errors were recorded for analysis using Kaltura screen capture. Tables were created in MS Word (See Appendix 6.7) that compare identification and categorisation results. Initial observations of the tables of results are given in Section 6.3.1. Descriptive statistics were used to calculate the raters' overall inter-rater agreement in these two areas.
- 2) The reasons behind any different categorisations were extracted from the transcriptions of the audio recordings made with Kaltura screen capture and the data from the semi-structured interviews. The transcriptions were analysed using thematic analysis.
- 3) Confidence scores were recorded by participants to show which errors or error types appeared to be more or less problematic for categorisation. These scores were transposed to a table for analysis.
- 4) Teachers' satisfaction with the depth of analysis provided by the LEA was informed by the data from the semi-structured interviews.
- 5) Possible uses of the results of LEA in the teachers' practice were informed by the data from the semi-structured interviews.

- 6) Areas for improvement to the guidance and framework were informed by all methods of data collection and analysis. The transcriptions were analysed using thematic analysis for these last three points

6.2.5.1 Research question 1) Is there an adequate level of inter-rater agreement when experienced language teachers are asked to identify LEs in a piece of student writing and categorise extracted LEs using the new framework?

Identification

To address the first part of RQ1, (identification), the four annotated documents (one from each pair of participants and one from myself) were analysed for similarity of identification of LEs.

37 LEs were identified by myself and numbered 1-37. The participants' annotated essays were examined to see if they had also identified them as such. For example, if, in my analysis, an error had been identified with the third word in the first line, this error was allocated a variable number of one. The other three scripts were checked to see if this had been identified as an error. If, for example, two of the three remaining scripts had identified it as such, that error would have been allocated a similarity score of 75% (three agreements divided by a possible maximum of four, multiplied by 100). The other three scripts were searched for LEs that I had not identified and the process was repeated for each error.

MS Excel was used to record all the errors that were identified and the number of times that they were identified. It was used to calculate the total number of errors and the number of errors identified by individual participants. The average agreement figure was calculated by adding up all the agreement percentages and dividing that number by the total number of different errors identified to give an overall score for error identification across the participants. The average of all these figures was expressed as an overall percentage similarity for error identification.

Categorisation

This section addresses the second part of RQ1 (categorisation).

Quantitative data analysis

Categorisation Scores

MS Excel was used to record the framework error codes that were allocated by each of the participants to the 37 extracted errors from their MS Word Docs (see Appendix 6.7). The copy and paste function was used to eliminate the possibility of clerical transposition error. The mode (the most common categorisation code) was found and was compared to the categorisation code allocated by myself to establish whether the framework was being used to categorise errors as I had envisaged.

An agreement score for each of the 37 extracted errors was calculated by taking the number of agreements between the seven participants (they were asked to record individual, not paired decisions) and dividing that number by seven, then multiplying the sum by 100. The percentages were added together and then divided by 37 to give an overall agreement figure for categorisation. For example, if all participants agree on the categorisation, it is allocated a score of 100%. If, for example, four of the seven agree on a categorisation, the percentage agreement figure is 57% (4 divided by seven, multiplied by 100). The mode was taken as a central answer. If a participant entered a different categorisation code, a dual categorisation code was generated for later analysis (see Section 6.3.1): these emergent codes were created where there was disagreement from some participants with the most popular categorisation choice (the mode). For example, with Error 4 - *Let *me discuss the topic*, five participants selected [B12 41 Formality error (Misselection)] (the mode), and two selected [B12 42 Formality (Blend)] an error code for this specific dual categorisation combination (CES) was created. For instances where there was even disagreement with no most popular categorisation choice, I decided which I felt was the most logical categorisation in line with the guidance, and error codes were created for disagreement against this.

Qualitative data analysis

The reasons behind the differences in categorisations were analysed using the transcripts of recordings made during the categorisation tasks to establish, where the participants mentioned them, their thinking behind their different categorisations. This produced further emergent codes when more issues were identified (all of these error codes are explained and quantified in Table 35 below).

Confidence scores

Confidence scores were also recorded for each error from each of the seven participants. The mean confidence score for each error was calculated to establish which of the 37 errors were categorised more confidently. This was also done for each participant to see if there were differences between individual participants. An overall single confidence score (for all participants and errors) was also calculated using MS Excel (See Appendix 6.7, Table 2) to ascertain how confident these participants were in categorising these errors using NewLEAF1.

6.2.5.2 Research question 2) What thoughts and opinions did participants have on the ease of use of the framework, and what problems did they encounter when using the framework?

To answer this question, the screen capture data and the semi-structured interview data were analysed using a thematic analysis approach.

6.2.5.3 Research questions 3 and 4) How satisfied are the teachers with the depth of analysis? And how might they use LEA results in their practice?

These questions were analysed using a thematic analysis of the semi-structured interview data.

6.2.5.4 Research question 5) Based on this information, how might the framework and guidance be improved?

This question was answered using the annotated scripts, the completed charts, the transcripts from the screen capture and the semi-structured interview data using thematic analysis.

Analysis of all recordings - Issues in analysing and coding data of recorded data

The recordings from both the think-aloud protocols and the follow up interviews were transcribed and a thematic analysis (Cohen *et al.*, 2018, Dörnyei, 2007) was used to analyse the data using Nvivo. Pre-conceived themes and emergent codes (Cohen *et al.*, 2018, Dörnyei, 2007) were used to group any issues that were identified during the analyses. The pre-conceived themes relate to the areas that I wanted to investigate. Emergent sub-themes in these areas, plus any emergent areas were coded and grouped. Azevedo *et al.* (2015:765) recommend transcribing, coding, recoding and then analysing the data. They also recommend

analysing the data in the context of the cognitive task being undertaken, which requires cognitive task analysis ‘in order to know the knowledge states, problem-solving operators, and problem-solving strategies (e.g., means-end analysis) that are to be used to segment individual statements’. The pre-conceived themes (and codes) of interest to the current study are:

- Issues in categorisation of LEs (Cat)
- Thoughts on ease of use (Ease)
- Thoughts on/satisfaction with depth of analysis provided by the results (do they think the results are suitably detailed for a student or teacher to act upon?) (Depth)
- How analysis findings could be used by teachers or with individual students and groups (Use)
- Comments to improve the guidance (ImpG)
- Comments to improve the framework (ImpF)

Issues and solutions relating to transcription and coding of LEA and interview transcripts

It is important that the development and testing of a coding scheme should be piloted (Cohen *et al.*, 2018, Dörnyei, 2007). Transcription requires careful planning to accurately transcribe, segment and then code the data. Decisions need to be made about what to do with: unintelligible parts; differentiating between verbal data and reading from the screen; segmenting (how does one determine the beginning and end of a segment) and coding.

To address these issues, the following solutions were implemented: The whole process was piloted two weeks before the data were collected by a researcher colleague at the university. He was asked to make a note of any issues he experienced. The codes were refined twice to avoid duplication and omission issues. Rather than guess at meaning, unintelligible parts of the recordings were represented with ellipses (...) so that inaccurate data did not enter the data set. Differentiating between reading from the screen and verbalising thoughts is not believed to be an issue as the text and the errors in the essays will be known to myself. Reading from the screen is shown in italics in the transcriptions and extracts in the results and discussion sections. (See Appendix 6.9 for an example of a recording transcript.) Segmenting issues were overcome by using Kaltura screen capture. As the screen is also video recorded, and shows the participants’

screens rather than the participants themselves, it was clear which LE (also numbered) was being discussed. Participants completed entries for each LE as they were discussed.

Regarding the coding of the data, in-depth comments relating to the pre-conceived themes above were sought, so emergent comments relating to the six areas above were coded. Data analysis was conducted by extracting the comments from the transcriptions and then creating and allocating codes to these themed areas, plus any others that emerge. Also, where participants categorised errors differently from each other, a code was created to show the dual categorisation potential when certain errors were categorised differently from the majority of participants. For example, if three participants categorised an error as type 1, two participants categorised it as type 2 and one participant categorised it as type 3, codes would be established for dual categorisation potential between type 1 and type 2, and also between type 1 and type 3 (see Section 6.3 for more details).

Observation of participants

Observation added useful data. When there were pauses in the discussions, it was when the participants were consulting the framework and guidance. This close reading of the framework and guidance was, at times, accompanied by other expressions or gestures of confusion or concentration, which demonstrated possible issues with the framework. Such non-verbal data was included in the transcripts in brackets (____).

6.3 Results and discussion

Results will be presented and discussed by research question. Extracts from the transcripts of the screen capture, observation and semi-structured interviews are used to supplement the quantitative data and present the themes (preconceived and emergent).

6.3.1 Research question 1 Is there an adequate level of inter-rater agreement when experienced language teachers are asked to identify LEs in a piece of student writing and categorise extracted LEs using the new framework?

Identification

To answer the first part of RQ1 (identification), as can be seen from Table 33 below, the overall inter-rater agreement score for the current study was 59.3%. Errors that were identified by all included easy to spot errors of form (e.g., Errors

2, 7, 13, 22, 25, 26, 31 and 34) and errors that clearly involve wrong word or phrase choices (e.g., Errors 32 and 33) or overly informal errors (Error 29). 59.3% can be considered quite low: too low if the framework is to be used to compare results of analyses completed by different analysts on the same compositions. The low figure could be mainly explained by the fact that I was quite strict in error identification, whereas the remaining participants were more lenient, particularly in the area of [Formality]: only I identified eight [Formality] errors as such (for example, Errors 1, 4, 10, 14, 16, 17, 19, 23). Formality error decision making is not a black and white decision, but there are shades of formality that would be more acceptable to one analyst more than another. Other errors that returned a low agreement score and therefore also contributed to the low overall score included instances of wrong near synonymous word or phrase choices (for example, Errors 3, 5, 11, 12, 18, 20, 24, 27, 42). It seems that for different analysts to identify LEs consistently, they must have similar tolerance of error levels and similar beliefs in what is and what is not acceptable, which would be a difficult objective for a guidance document or training programme. However, it must also be stated that all testing/scoring systems are variable to the user to an extent, which is why there are moderation meetings when group marking, for example.

No	Error	R	P1+2	P3+4	P5+6	% agr
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	1	0	0	0	25
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	1	1	1	1	100
3	perfect academic performan is a <i>basic quality</i> that every student wants.	1	0	0	0	25
4	Let <i>me</i> discuss the topic	1	0	0	0	25
5	and indicate how I disagree with the <i>topic</i> .	1	0	1	0	50
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	1	0	1	0	50
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	1	1	1	1	100
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	1	1	1	0	75
9	As a student who <i>finishes</i> amount of tasks and homework,	1	1	1	0	75
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	1	0	0	0	25
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .	1	0	1	0	50
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .	1	0	1	0	50
13	different kinds of books <i>essaies</i> , files and so on.	1	1	1	1	100
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .	1	0	0	0	25
15	an essay must be finished <i>in</i> the end of a semester.	1	1	1	0	75
16	<i>Is it the most important ?</i> I do not think so.	1	0	0	0	25
17	Is it the most important ? <i>I do not think so</i> .	1	0	0	0	25
18	So, what score can a student obtain is determined by <i>how to organise</i> words to indicate points	1	0	0	1	50
19	So, <i>in my opinion</i>	1	0	0	0	25
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	1	0	1	0	50
21	to <i>gain great academic performance</i>	1	1	1	0	75

No	Error	R	P1+2	P3+4	P5+6	% agr
22	, the listening skill is <i>a prior skill</i> for a students.	1	1	1	1	100
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	1	0	0	0	25
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	1	0	0	0	25
25	content of what proffesser says clearly.	1	1	1	1	100
26	content of what proffesser says clearly.	1	1	1	1	100
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	1	0	1	0	50
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	1	1	1	0	75
29	Who <i>wannna</i> be a silent person ?	1	1	1	1	100
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills	1	1	0	1	75
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	1	1	1	1	100
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	1	1	1	1	100
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	1	1	1	1	100
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	1	1	1	1	100
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .	1	1	1	0	75
36	I think, the most important is not only reading skill. <i>Other skills are not. either</i> .	1	0	0	0	25
37	the best way to <i>obtain the biggest</i> progress	1	1	1	0	75
38	Some people think the <u>Reading</u> skill is the most important	0	1	1	1	75
39	<u>Thus</u> , how to read them fast and correctly is a fundamental ability	0	0	1	0	25
40	is determined by how to organise words to <u>indicate</u> points.	0	0	1	1	50
41	Then, the listening <u>skill</u> is a	0	0	1	0	25
42	So, listening is a very effective way to <u>obtain</u> new	0	0	1	0	25

43	So, listening is a very effective way to obtain new <u>things</u> .	0	0	1	0	25
Totals		37	20	30	15	59.30

Table 32 Total number of errors identified by different participants

Categorisation

To answer the second part of RQ1 (categorisation), results are analysed and discussions of the different categorisations of each of the 37 errors are provided. Table 2 in Appendix 6.7 shows the categorisation decisions and associated confidence scores. Table 34 below shows several interesting points about the results of the analysis and the inter-rater agreement when the framework is used: there was a minority of cases where all analysts were in total agreement over the precise type of error (13 out of 37 errors, or 35%). These were errors in [A3 Spelling], [A4 Two words should be one], [B3 Wrong Word (Sense relations error)], [B4 Preposition] and [B12 Formality]. This percentage is too low. Whilst 100% may be too much to hope for, a figure of 80% would have been more acceptable. There were eight instances when an error was categorised in four ways (21% of all errors) and 10 instances where the same error was categorised in three different ways (27%). For example, in nine instances there was disagreement between [Field error] and [Sense relations error]. This was the most frequent disagreement, accounting for 8% of all errors (see Table 2 in Appendix 6.7). Again, these figures are not acceptable and need to be improved upon if the framework is to show consistent use between raters. From this information alone, it is clear that much work is needed to be done in the standardisation of or alterations to the guidance and framework to improve inter-rater agreement.

Agreement in rater categorisation	Instances of agreement	Percentage of all errors
1 (Total Agreement)	13	35
2 (Partial agreement)	6	16
3 (Some agreement)	10	27
4 (Little agreement)	8	21
Totals	37	100

Table 33 Instances of rater categorisation agreement by different participants

Confidence in categorisations

69% of confidence scores were given as 3s. Table 2 in Appendix 6.7 also shows that overall, participants were happy with their categorisations. Although 69%

could be seen as satisfactory for the first version of the framework, I felt that this figure could be improved. The average participant confidence score for all categorisations was 2.7, out of a maximum of 3, which suggests that participants overall felt very confident that they were allocating errors to categories correctly. This demonstrates that they feel that they can use the framework and guidance to correctly categorise extracted errors to the correct categories, but this does not indicate sound agreement. Individual participants registered an average confidence score of between 2.37 and 3, with an average confidence score of 2.7 per error categorisation. Unsurprisingly, most 3s were recorded by myself, the author of the framework, as I felt I understood it thoroughly, and the lowest by the least experienced participant, who also demonstrated the least familiarity with some of the terminology used in the framework and guidance.

The decision to pair analysts during the data collection phase ensured that they had to discuss and voice their thoughts before they entered their individual error codes into the chart. Whilst there were data collection advantages for adopting this approach, it often resulted in analysts swaying each other's opinions. There may have been different results if they had completed the analysis independently. At times they agreed with each other when one was helped to understand the co-text, framework and guidance. This occurred frequently, even though they were asked to record their decisions independently following discussions. It can be seen that there were only six instances (16% of all errors) where one of the pair of analysts' codings were different from the other's. The pairing of participants can be seen as both a problem for subjective data collection and a benefit, as pairing requires participants to voice their thoughts and check that both are following analysis guidelines.

As can be seen from Table 2 in Appendix 6.7, the overall inter-rater agreement score is 67%. Possible reasons for this rather less than satisfactory percentage score can be found in the next section. Given the importance of producing a framework that will generate consistent results for teachers and researchers in different contexts, this figure can be seen as too low and changes must be made to either the framework or the training supplied to use it, or both.

Further statistical reporting for individual errors can be found in the section below.

Agreement issues with specific errors

In this section, the different ways that the 37 errors were categorised by the seven participants are presented and discussed to identify issues related to dual categorisation. A list of suggested actions for NewLEAF2 are presented in Section 6.4.1.

Summary of categorisation issues and codes

With regard to issues in the categorisation of the 37 extracted errors, Table 35 below shows the number of instances of emergent micro codes from the analysis. 37 different issues were identified. It can be seen that the most common cause of confusion was lack of understanding of the framework or guidance (11 instances, accounting for 10% of all issues). These could have been issues with the clarity of the documents or the participants' understanding of the documents. The next most common was dual categorisation potential between [Sense relations] and [Field] errors. This could, perhaps, have been predicted as an issue of subjectivity. There was a large number of categorisation issues that yielded only one instance of its type.

Themes	Emergent code	Count	% of all issues
Issues in categorisation of LEs			
Lack of understanding/clarity of framework or guidance	LUFG	11	10
Dual categorisation potential between Sense relations and Field errors	CSF	9	8
Ensure participants consult full text to understand meaning of errors.	PCT	8	7
Understanding terminology Issue	UTI	8	7
Dual categorisation potential between One word and Paraphrase	COP	8	7
Reduce the amount of categories or information in the guidance	RAI	7	7
Dual categorisation potential between the same error type but different surface taxonomy	CES	6	6
Dual categorisation potential between Coherence and Paraphrase	CCP	5	5
Dual categorisation potential between Paraphrase and Coherence	CPC	4	4
Dual categorisation potential between Sense Relations and Paraphrase	CSP	3	3

Different categorisations caused by differences in reformulation of errors.	DRD	3	3
Dual categorisation potential between Cohesion and Sense relations	CCS	3	3
Dual categorisation potential between Formality and Wrong word	CFW	3	3
Dual categorisation potential caused by decontextualized errors	CDE	2	2
Uncertainty over whether incorrect register constitutes a LE.	ILEF	2	2
Dual categorisation potential between Paraphrase and Connotative meaning	CPC	2	2
Dual categorisation potential between Underspecification and Paraphrase	CUP	2	2
Dual categorisation potential between Coherence and Underspecification	CCU	2	2
Dual categorisation potential between Cohesion and Verbosity	CCV	1	1
Dual categorisation potential between Formality and Paraphrase	CFP	1	1
Dual categorisation potential between Formality and Sense relations	CFS	1	1
Dual categorisation potential between Coherence and Wrong word	CCW	1	1
Dual categorisation potential between Coherence and Suffix	CCS	1	1
Dual categorisation potential between Sense relations and Connotative meaning	CSC	1	1
Dual categorisation potential between Sense relations and Verbosity	CSV	1	1
Dual categorisation potential between Verbosity and Suffix	CVS	1	1
Dual categorisation potential between Verbosity and Cohesion	CVC	1	1
Dual categorisation potential between Formality and Coherence	CFC	1	1
Dual categorisation potential between Formality and underspecification	CFU	1	1
Dual categorisation potential between Formality and Connotative meaning	CFCM	1	1
Dual categorisation potential between Formality and Miscellaneous	CFM	1	1
Dual categorisation potential between Paraphrase and Cohesion	CPCo	1	1
Dual categorisation potential between Sense relations and Preposition	CSPr	1	1
Dual categorisation potential between Connotative meaning and Coherence	CCC	1	1

Dual categorisation potential between Suffix and Paraphrase	CSPa	1	1
Dual categorisation potential between Paraphrase and Arbitrary combinations	CPA	1	1
Differences of opinion of acceptable formality	DAF	1	1
Totals		107	100

Table 34 Quantitative analysis of emergent codes for categorisation issues

To summarise responses to RQ1: there was much dissimilarity when experienced language teachers were asked to identify LEs in a learner's writing (only 59% agreement) and further insufficient similarity (only 67% agreement) when they were asked to categorise 37 extracted LEs using the new framework. Following on from the data analysis above, if a better degree of agreement is to be achieved, significant changes are to be made to the framework itself and the guidance or training used to familiarise the participants to use the framework.

During the interviews, participants were asked specific questions (see Appendix 6.8 for interview schedule) relating to research questions 2, 3, 4 and 5. The summary totals of emergent codes relating to each research question are given at the foot of each of the next three relevant sections.

6.3.2 Research question 2) What thoughts and opinions did participants have on the ease of use of the framework and what problems did they encounter when using the framework?

Discussion here is grouped under the pre-conceived themes of clarity of guidance to use the framework, clarity of the different categories, clarity of differences between the categories and overall ease of use of the framework. Emergent themes are presented in Table 36 below with their frequency.

Clarity of guidance to use the framework

Participants were asked 'How clear to you was the guidance to use the framework?' and were given the following possible Likert scale responses: Completely clear, Quite clear, Not sure, A little confusing, Very confusing. All interviewees mentioned that the guidance was 'Quite clear. However, this does not account for the fact that there was much difference in the application of the guidance in the identification and categorisation tasks. When asked to elaborate on their answers, Participant 2 stated that separating lexical and grammatical

errors was helped by the guidance, but there were still instances when she was unsure. She also mentioned that she found the terminology in the guidance a little difficult and she thought that 'not all TESOL teachers would be able to cope with it'. She admitted having to look up the meaning of one item and struggled to recollect some of the reading in linguistics she had done, as she doesn't habitually use these terms. However, she also mentioned that 'certainly when we were using the framework, it became easier towards the end' – i.e., once she had had some practice with it.

It seems that even highly trained and experienced language teachers would benefit from a more simplified set of instructions for guidance. The framework itself could be improved by being simplified. However, P2 also mentioned that as she became used to the framework, it became easier to use. Perhaps it might have been better to measure agreement between participants after they had become familiarised with the framework.

Clarity of the different categories

Participants were asked 'Are the categories in the framework, i.e., the types of error to be allocated to them, sufficiently clear for you?' They were given the following Likert scale responses: Completely clear, Quite clear, Not sure, A little confusing and Very confusing. All mentioned that these were quite clear, but again, there were differing opinions on how to categorise many of the errors. When asked to elaborate on their answers, P2 mentioned that the guidance and examples 'certainly helped'. Participant 3 stated that there was some ambiguity over [Connotative meaning]. The examples given (famous/notorious) 'could be considered as gradable adjectives and there could be another category for these'. She also added that it would be very difficult to have 100% mutually exclusive categories. Participant 5 stated that there was a lot of metalanguage, making it 'quite heavy'. There were issues with [Connotative meaning] and [Formality] and 'ambiguity between the two'. P2 stated that most of them were clear. Having the supporting document that explained what [Irreversible binomials] were with examples helped, as did those that explained [Coherence] and [Cohesion] as 'sometimes I find it difficult to differentiate between those two'. She found it easier to decide which vertical column to allocate to than the horizontal (surface level taxonomy). She felt that this was very clear.

P3 Mentioned that the categorisations 'could be seen in different ways', but also that having a system was very useful. P5 mentioned that it would be better if there were 'more examples'. He mentioned often thinking that an error could fall into more than one category. He suggested including [L1 interference errors], but then stated that this could come later. He again highlighted the possibility of categorising an error as [Connotative] or [Formality] or [Paraphrase] error. He mentioned it was more these 'socio errors' (usage errors) that proved more troublesome than the surface errors that he is more used to dealing with.

Clarity of differences between the categories

Participants were asked whether the differences between the categories were clear enough to use the framework easily, and were offered the following Likert scale responses: Completely clear, Quite clear, Not sure, A little confusing and Very confusing. All participants answered, 'Quite clear'. When asked to explain their answers, P2 stated that dual categorisation occurred quite often, and discussing and completing the task with a colleague 'helped to identify that'. She mentioned having some difficulty with error repair: when one considers what needs to be changed to repair one error, it does not fit with the co-text of the rest of the sentence and another error becomes evident, such as a grammatical one. This seems to be an issue of error counting. However, she also stated that having the error identified for her in italics also helped. She also mentioned that discussing intended meaning with the learner would help, which reflects Corder's (1973) suggestion. She also mentioned that this is time consuming and not always possible. P2, when outlining the difficulty she had with error repair, seemed to be focussing on causality, which, due to its speculative nature is probably best avoided at the categorisation stage. She also used examples of grammatical errors. However, this may still be an issue with LEs, so perhaps in the guidance, Lennon's (1991) concepts of domain and extent need to be incorporated and it needs to be made clearer that causality should be ignored for this stage and the error categorised, not the possible reformulation. P3 stated that the framework does not need any more categories, as this may 'overload the practitioner' when conducting analysis. She stated that there will always be a need for a miscellaneous category and did not think that a perfect framework was possible. However, she also stated that the framework would help the non-linguist

practitioner to analyse LEs. She felt the framework was 'good enough'. P5 mentioned that the metalanguage was 'quite heavy' again and 'took quite a bit of time to work out where to put the error'. He repeated the issue of frequent dual categorisation, but then politely put this down to his own inexperience of analysing errors in such depth, rather than his experience of using the correction code symbols that he uses with his current learners (he mentioned that there are five or six codes that he would use). 'I'd band all of these error types into one generic code'. He also mentioned that with more familiarity, he might not struggle as much.

Overall ease of use of the framework

Participants were asked 'Overall, how easy or difficult was it for you to use the framework?' They were given the following Likert scale responses: Very easy, quite easy, Not sure, A little difficult or Very difficult. P2 stated 'Not sure' (because it started as difficult, but ended up as easy). P3 and 5 stated 'Quite easy'. When asked to explain their answers, P2 mentioned that it was a little difficult because it was the first time she had used it but it became 'quite easy towards the end, with practice or familiarity.' 'If I had been using it on my own, it would have been easier as I wouldn't have to rationalise my choices all the time, however, verbalising your thinking does help you to slow down and ponder it.' She expressed difficulty in differentiating between [Sense relations] and [Field error] difficult at the beginning, but this became simple towards the end. P3 stated that she rated the ease of use as 'quite easy' due to the issues of dual-categorisation ('slight' hesitations and slight interpretation issues). She also added that it seemed 'well-thought through'. She mentioned that some of the points in the guidance for identifying and separating out grammatical errors were bundled together, whereas others were not. More consistency could be achieved here in presentation. P5 stated that there was a lot of information for the marker to read before allocating errors. He also mentioned a lack of familiarity with the framework again, which made it take longer. Interestingly, P5, during the analysis, did not seem to be taking more time than the other participants, nor did this time seem unduly long. Also, working with two axes made it seem longer to him, which he says he found 'quite a challenge'.

Themes from semi-structured interviews	Emergent code	Count
Issues in dual categorisation of LEs	CLE	5
Discussion with partner helped with dual categorisation issues.	DDC	5
Students would benefit from this	SBT	3
The categories were 'quite clear'	CQC	3
Differences between categories were 'quite clear'	DBC	3
Uncertainty over whether incorrect register constitutes a LE.	ILEF	2
Ambiguity over meaning of [Paraphrase]	AMP	2
Ambiguity over meaning of [Connotative]	ACN	2
Ambiguity over meaning of [Formality]	AFO	1
Ambiguity over meaning of [Coherence]	ACC	1
Ambiguity over meaning of [Cohesion]	ACS	1
Ambiguity over meaning of [Arbitrary combinations]	AAC	1
Ambiguity between [Field] and [Sense relations] errors	AFS	1
Some grammatical points bundled together in guidance	GPB	1
Too much information	TMI	1
Working with two axes 'quite a challenge'	TAQ	1
Powerful tool	PTO	1
More categories than I am used to dealing with	MCI	1
Time consuming to use	TCU	1
Verbalising thoughts to a partner helps with categorisation thought process	VTH	1
[Paraphrase] provided a 'catch all'	PPC	1
It is a systematised method	ISM	1
Bring in a 'gradability' category	BGC	1
Dual axes provide more depth	DAP	1
Students would appreciate it	SWA	1
Doubt about whether a 100% mutually exclusive taxonomy of LEs could be produced	DMT	1
Clarity of categories is fine	CCF	1

Framework would act as a good linguistic training tool for teachers	FGL	1
Italicising the error helps with categorisation	IEC	1
Issues with Error Repair	IER	1
Support for learner consultation over intended meaning	SLC	1
Rejection of learner consultation over intended meaning	RLC	1
Issues in distinguishing between lexical and grammatical errors	LGE	1

Table 35 Themes extracted from the transcription that relate to ease of use

6.3.3 Research question 3) How satisfied are the teachers with the depth of analysis?

P2 stated that some categories were used more than others. Others could be lumped together (e.g., [Suffix] and [Prefix]) to simplify the framework. She mentioned that she avoided the ‘temptation to use the miscellaneous category’ when they were uncertain over which category to use. The [Paraphrase] category also provided a possible ‘catch all’ option and she wanted to avoid this category so that she could allocate to the more specific category, where possible. It seems P2 may have misunderstood the intended meaning of [Paraphrase]. It was meant as ‘this erroneous phrase contains errors that span across more than one word and the phrase needs to be re-written’, but perhaps P2 understood it as the *learner* had paraphrased something incorrectly. The guidance and framework need clarification here. She stated that the framework is ‘more detailed than the standard error correction codes that we use. For example, WW (wrong word) or WF (wrong form). This can be frustrating for students as, in my experience, they may not know what the right word is’. The LEA framework ‘might help’ narrow the error down for students, but P2 also stated that the terminology would also confuse the learners. She mentioned that it should come with an explanation of the benefits to learners in the same way that she always explains to her learners why she currently uses error correction codes. She also stated that the results are more detailed than the error codes she is using, and they are better explained.

P3 commented that ‘It’s good to have this level of depth. Also, this could act as a developmental linguistic training tool for teachers. It’s a real positive. But this level of depth shouldn’t be ‘offloaded’ onto the learners with feedback. It wouldn’t make sense to them and would provide too much feedback to the learners, but to clarify

for the marker what has gone wrong is very important'. One may need to distinguish between the usefulness for the teacher and the usefulness for the student. 'So used with discretion by the teachers, it will be pretty powerful.' She added that lexical gradeability errors could be brought in.

P5 stated that colleagues may feel that there may be too much detail, rather than too little. 'It's 'heavier than what I'm accustomed to in terms of metalanguage and number of categories.' He mentioned it may be too much for teachers to do on top of a busy teaching schedule. However, this could be a criticism of the teacher's workload, not the framework itself. P5 also mentioned that when he was doing the task, he felt that in his practice, he could offer his students more depth of analysis of their errors, but busy teachers will find it easier to use a simplified set of error correction codes. He questioned the value of his own set of codes to the learner, so 'something like this would definitely benefit the TESOL practitioner and the learners. I guess it's a case of finding the happy medium between what teachers do at the moment and something more in-depth'.

The emergent codes relating to depth of analysis are summarised below in Table 37.

Qualitative comment	Emergent Code	Count
Too much depth is provided by the analysis	TMD	3
A good level of depth is provided by the analysis	TLD	1
Framework will provide more depth of feedback than current system of error reporting	FPD	1

Table 36 Themes extracted from the transcription that relate to depth of analysis

6.3.4 Research question 4) How might they use LEA results in their practice?

P2 stated that she probably wouldn't use the LEA framework each time she marked a set of essays, but using it occasionally on a cohort's work would help her to identify errors that occurred frequently with many students, which would lead to useful remedial teaching. However, not all categories would lend themselves to a whole lesson on remedial teaching: spelling, for example, would not.

P3 stated that the framework would be useful to focus on the type of errors that a whole class makes, and that it may be impractical to focus on all the errors that each individual made: this would be 'too much work for everybody'. She also stated that a 'teacher using this would be able to spot error patterns among the group' and it would be 'particularly useful for one-to-one teachers, who might be grateful for a systematised method for analysing errors, and this would be useful for a basis for discussion, but still discretion and professional judgement will be important. Perhaps a guide or accompanying handbook would be useful, something more attractive with colour and illustrations. Otherwise, this could be heavy, as some practitioners may not have the same understanding of linguistics.'

P5 also stated that the framework could be used as it analyses errors in much more depth than the correction symbols that are used to 'find out more about your learners', as there is also the surface level taxonomy, which might enable the teacher to identify patterns of errors. This may then inform the teacher of whether there were L1 issues, which may help inform the teaching in future classes. He also stated that if learners received some training, then results could be passed over to them and they would 'greatly benefit and appreciate it because there would be more attention on their errors'.

Of course, one could use error codes from the LEA framework as one would an existing set of error correction when marking student work. For instance, when marking a composition, one could, instead of copying and pasting an error into the corresponding framework cell or tallying the number of instances of that error in that cell, simply write, for example, '1' next to where a student has omitted a suffix.

Table 38 below summarises the emerging codes relating to how LEA data could be used in teaching practice.

How analysis findings could be used	Emergent Code	Count
Probably wouldn't use it each time she marked essays	PWU	2
Helpful to identify commonly occurring errors, which would lead to remedial teaching	HIC	2
Useful to find out more about learners' error patterns	UFM	2
Useful to focus on type of errors that a whole class would make	UFE	1
Useful for one-to-one teaching	UFO	1
May inform teacher of L1 interference errors	ITL	1
Share results with learners (with some training)	SRL	1

Table 37 Themes extracted from the transcription that relate to how analysis findings could be used in their practice

6.3.5 Research question 5) Based on this information, how might the framework and guidance be improved?

6.3.5.1 Analysis of results and recordings of identification and categorisation tasks

The analysis, as described in Section 6.2.5, provided several suggestions for improvement to the framework and guidance. These are summarised in Section 6.4, below.

6.3.5.2 Analysis of semi-structured interview data

The semi-structured interviews revealed some useful information for the development of the framework and guidance.

When asked whether they had any general comments on the guidance or framework, P2 mentioned that training for teachers would be required (some norming) to make it accessible and to overcome issues with terminology and more examples, which she found useful in understanding the differences between categories. 'Perhaps a self-training quiz to check they've got it right before they use it live with students. Without the training, they may revert to the miscellaneous category or simply go back to using wrong word or wrong form type error correction symbols that they might have been using before.'

Participant 2 stated that the framework would be easier to use alone than with a partner. Perhaps she felt that she had to explain the framework to the partner too much or that the discussion slowed the whole process down somewhat.

P5 mentioned that there would need to be some training for teachers to use the framework during induction for teaching on a course. He felt that how errors are dealt with is sadly neglected in teacher training. He felt that this would be really beneficial 'do the teacher training and then do the learner training as well, but we don't have something clear and structured in place at the moment'.

P3 suggested user friendly and well-thought through teacher support material with examples and comments about how deeply to use the linguistic metalanguage and suggestions for different types of students and a helpful warning about throwing all the errors back at the learners.

P5 stated that there will probably always be a need for a miscellaneous category. This is true as it could be the case that in future essays an error type emerges that was not encountered in the data used in these studies. Also, this category would be important to retain as it acts to ensure all errors are captured in some way, as it could be the case that analysts could not confidently allocate an error to a category, so they may omit it from the analysis altogether.

Table 39 below summarises the emergent codes relating to improvement of the guidance and framework.

Comments to improve the guidance	Emergent Code	Count
Too much metalanguage	TMM	6
Guidance was 'quite clear'	GQC	3
Need more examples	NME	3
Examples were helpful	EWH	2
Provide training/norming for teachers	PTT	2
A more colourful, user-friendly guide	MCG	2
Guidance helpful to distinguish between LEs and GEs	GHD	1

Comments to improve the guidance	Emergent Code	Count
Provide convincing argument of the benefits	PCA	1
Comments to improve the framework		
Framework became easier with use	FEU	4
More depth than existing error correction system	MDE	4
Include L1 interference category	ILI	1
Easy to allocate to surface level taxonomy vertical columns than error type horizontal ones	EVH	1
It's 'good enough'.	IGE	1
There will always be a need for a miscellaneous category	NMC	1
Helpful for the non-linguist practitioner	HNP	1
Better explained than existing error correction system	BEE	1
Overall ease of use – not sure	OEN	1
Overall ease of use – quite easy	OEQ	1
Easier to use with fewer categories	EUF	1
Easier to use alone than with a partner	EUA	1
Combine some categories	CSC	1
One version for teachers and another for feedback to students	OVT	1

Table 38 Themes extracted from the transcription that relate to how the guidance and framework could be improved

6.4 Summary of recommendations for refinement, based on the results and discussion

This section presents a list of suggestions for the development of the LEA framework and guidance: firstly, ones to be adopted and then ones not to be adopted. Rationales are offered in all cases.

6.4.1 Suggestions to be adopted

6.4.1.1 Clarify terminology in the framework and guidance

- Despite their relatively high level of qualifications, there were several instances of participants not understanding the terminology used in the guidance and framework. It is logical to believe that if these participants, given their position and expertise, struggle with the terminology, other language teachers are almost certain to do so. If inter-rater agreement is to be achieved, the participants must be able to understand the accompanying documentation. Therefore, wherever possible, clear, jargon-free explanation is to be utilised. Therefore, the framework and guidance should be simplified wherever possible. However, this inevitably leads to the use of more words to explain concepts that could be defined with a single word at times.
- Clarify explanation of [Arbitrary combinations and binominals]. Use [Misordering of words in fixed phrases] instead.
- Rephrase [Conjunctions] to [Linking devices] in guidance.
- Improve definition of [Connotation] or remove category. It was decided to remove this category, as the concept of connotation was quite subjective, and there is scope for dual-categorisation with [wrong word]. Having separate [Wrong word (word class)] and [Wrong phrase] categories will cover this and stop dual categorisation possibility.
- Clarify wording of [Paraphrase] category to [Errors across multi-word units or phrases] as one participant interpreted this as ‘the student has paraphrased something incorrectly.’

6.4.1.2 Remove ambiguity in framework and guidance

- Remove category 42 as it is hard to imagine a [Formality blend error], where students incorrectly blend together two incorrect levels of formality within a genre of writing.
- Delete reference to pronoun use in [B5] as this conflicts with advice for identification of grammar errors. Give [Pronoun use] its own category. Some may feel this is more of a grammar word, but it is included here, as it

carries lexical meaning and its inclusion provides a wider coverage of possible errors, therefore enriching the framework. Therefore, remove reference to pronoun use in the guidance where it explicitly states in the identification section that pronouns should be considered grammatical and excluded from the analysis.

- To ensure better inter-rater error identification, in the guidance, advise that even errors that they would normally 'let pass' should be included in the analysis. This would help to avoid some subjectivity in error identification.

6.4.1.3 Reduce potential for dual categorisation

- Categorising single word errors by word class would provide a series of mutually exclusive categories in the taxonomy, and would be more familiar to English Language teachers, whereas [Sense relations] proved hard to understand.
- Merge the categories for [Sense relations] and [Field] error to form one category for [Wrong single word] and one for [Wrong phrase], as this caused the most instances of dual categorisation. It, at times, is not a matter of an error being one or the other in the former pairing, but perhaps errors in these areas sit on a scale from one to the other and decisions made here can be subjective. This is a pity, as the categories of [Near synonym] and [Field error] feature in previous frameworks quite often and errors in these areas would be useful to report to students.
- Clarify that the [Wrong word] category is for single vocabulary items and multi-word items is for whole, or parts of, [Phrase] errors.
- Include lexical substitution errors as part of the wrong word class categories to avoid dual categorisation.
- Make [Underspecification (more details required)] a sub-category of [Phrase] error. [Single missing word] category would be part of a word class taxonomy.
- Make [Conjunction] errors a sub-category of [Wrong word].
- Clarify that [Coherence] is only for when you have tried to understand by re-reading, but cannot understand the phrase or word (which does not exist in English).

- Have [Formality] errors as a sub-category of [Phrase]. When a single word error with [Formality] is made, it would also be categorised under word class. This would avoid dual categorisation problems.
- Remove [Miscellaneous] category to force participants to make a more specific selection. This could be replaced by an [Uncategorisable error] category, which will act as a place or 'bin' to ensure that all errors are recorded and will also allow for future framework refinements. However, it is not currently envisaged that this will be used in a final version.
- Include [Irreversible binomials and trinomials] as a subcategory of [Phrase error (misordering in fixed phrases)].
- [B7 Wrong Words (Statistically weighted preferences)] caused some terminology confusion and seems overly specific and potentially quite a rare error. There is also potential for dual-categorisation with [Wrong word]. Therefore, the former should be removed.
- Remove the [Cohesion] category. Not only was this term misunderstood and overly-used, but there is also potential for dual-categorisation with [Wrong word] or [Phrase].

6.4.1.4 Provide clearer advice for categorisation in the guidance

- Clarify that infinitive 'to' is a grammatical error and should be excluded from the framework.
- In the guidance, ask participants to only consider the current utterance. Make it clear that the categorisation is for the error that they see, not the reformulation that is required.
- Advise to avoid allocation based on perceived causality. This would be speculative (Zimmermann, 1986).
- Ensure errors are presented in context and that analysts try to make an expert-like, plausible reconstruction of the error.
- Ensure the full composition is available and referred to by analysts. It was originally thought that providing the full sentence that contains the error would be sufficient to provide context to understand the intended meaning,

but this proved not to be the case. This shows that meaning of an utterance is found, not just in the co-text, but in the wider text also.

- Encourage analysts to come up with an expert-like, plausible reconstruction and if they cannot, allocate to coherence category.
- Reduce the amount of information in the guidance to avoid confusion. This has been achieved by taking the examples from the guidance and including them in the framework. Now there is advice for analysis in the guidance and examples in the framework only.
- Clarify what a [Blend] is. One participant thought that it was a blend of incorrect items.
- Clarify that [Coherence] is only for non-existent words or phrases they simply do not understand.

The suggestions above were implemented to redesign the guidance and framework. NewLEAF2 can be found in Appendix 6.1.

6.4.2 Suggestions not adopted

6.4.2.1 Remove ambiguity in framework and guidance

- Clarify what is meant by lexical substitution. This change would be unnecessary as errors in lexical substitution would be categorised under [Wrong word] in the new version.
- Reconsider the order of categories. If, when there is the possibility of dual categorisation, the advice in the guidance is to allocate to the more specific category towards the top of the section in the framework. I should ensure that the more specific categories are indeed at the tops of each of the two sections. This instruction seemed to cause some confusion as it may be subjective as to which category was the more specific or serious error type. Therefore, it was not incorporated into the new guidance.
- Clarify the difference between [Paraphrase] and [Wrong sense relations phrase]. This was not adopted as the [Sense relations] category has been removed. However, the [Paraphrase] category title has been re-phrased so that it is not a description of what needs to be done to make meaning clearer, but better describes the type of error made: [Phrase].

- Reduce the number of categories in the framework. This was suggested by two of the participants. However, the total number of categories has not been reduced so that depth of analysis can be achieved.

6.4.3 Other observations

It seems that despite the large number of dual coding issues and resulting codes, very often participants were thinking along very similar lines, but differed for the reasons laid out in Section 6.3.1 (they often agreed on the error type, but differed on the surface taxonomy level, or they entered different numerical codes, but largely agreed on the main typology of error. Therefore, having fewer main categories with some sub-categories and looking at the reliability of codings within the main categories only would show greater inter-rater agreement).

Undoubtedly, participants were assisted by the fact that the error that they were asked to focus on was in italics. It is feasible to think that if the specific error had not been in italics, participants could have interpreted and therefore categorised the utterance as a different kind of error.

Some participants were better at understanding what was meant by the student than others. This resulted in different understandings of errors and therefore different categorisations.

Perhaps some participants did not fully read and understand the framework and guidance before they started the analysis (Participants 3 and 4 only noticed the [Formality] category after discussing the categorisation of an error after a few minutes). This would inevitably lead to a reduction in agreement.

Very often, during the error analysis, analysts instinctively followed Corder's (1973) recommendation that a plausible reconstruction of the error should be made before allocating to category of error. However, different analysts, understandably, reconstructed erroneous utterances differently, and consequently arrived at different categorisations, depending on the various reconstructions. This suggests that the most likely plausible reconstructions should be explicitly encouraged to help analysts understand the correct meaning and categorise the actual error, not the reconstruction.

These points, and others, suggest that 100% agreement may be an unattainable target and perhaps the best that one could hope for would be to create a framework that is as reliable as it can be, given the subjective nature of error repair and different interpretations of intended meanings.

6.5 Conclusion

This section will provide a summary of the study, examine the limitations of the methodology used and the issues with the new framework

6.5.1 Summary

This study tested the new framework on practising, highly-qualified and highly-experienced English Language Teachers. Three pairs of teachers were asked to identify LEs in a student essay and then categorise 37 extracted LEs using the new framework. They were observed whilst doing so and subsequently, three individuals were interviewed to ascertain their thoughts on how easy it was for them to use, their thoughts on the depth of analysis that the results offered, what they might do with the results in their own practice and whether they had any suggestions to improve the framework and guidance. The participants stated that they found the framework easy to use and had some suggestions for how the results might be employed. Despite the fact that the framework was previously found to work well for the designer (see previous study), when used by others, there was great variation in the errors identified and there were far too many different ways that many of the identified errors could be categorised by different participants. This raised serious doubts about the inter-rater agreement of the framework. The guidance was also found to lack clarity in places, often due to the amount of information involved and confusing meta-language that was used. Data from the identification and categorisation tasks and the semi-structured interviews were used to redesign the framework. This included a new approach which separated the semantic Part B of the framework more clearly into errors with single words and errors with phrases. The former incorporated a word class categorisation approach as this would lead to greater inter-rater agreement.

6.5.2 Limitations of current study

This section deals firstly with methodological issues used during the study and then with issues relating to the framework itself.

6.5.2.1 Methodological limitations

Results using only six participants would not have been significant, but the participants were used again in Study 5, in the next chapter, for variable control purposes. Results would be useful to measure changes in agreement scores when comparing NewLEAF1 and 2.

Azevedo *et al.* (2015) recommend training the participants to think aloud by playing them a video of a subject performing a similar task whilst thinking aloud. They also suggest providing more than one video, showing a different type of task being considered and allowing the subjects to practice before completing the task at hand. This was not done for logistical reasons: mainly use of the participants' time. However, it became clear from examining the transcripts that some participants had not fully understood the guidance or framework, or read the written instructions before doing the tasks. Therefore, it would be sensible to provide instructions/presentation of the guidance and framework before analysts perform LEA for the first time.

The participants' thinking behind their LEA choices was sought using think aloud in pairs. Despite the fact that they were asked to record their own individual choices, identifying and categorising LEs in pairs undoubtedly swayed participants' choices in these areas. However, their think aloud decision-making process during the two tasks was sought and the methodology provided much useful and rich data to refine the framework and guidance. Furthermore, the discussion helped them to understand the guidance.

One issue with the process of analysing the identification or error data was that different participants identified the extent of the same errors in different ways. For example, some identified and underlined an error as follows 'a very crucial factor' whilst others identified and underlined the error as follows, 'a very crucial factor'. This was an issue with error counting, as reported by Lennon (1991). A decision was taken to view these as the same error and apply this rule to any errors where

the underlining overlapped. This could be seen as imperfect methodology, but if each error where the extent of the error differed was recorded as a separate error, the inter-rater agreement score would have decreased quite substantially, giving an inaccurate picture of inter-rater agreement in this area.

6.5.2.2 Issue with the framework

The framework no longer has a [Collocation] category. This is a shame given the importance of multi-word unit teaching in modern methodology. However, given the ubiquitous nature of collocations, inclusion of such a category would have caused far too many dual categorisation issues. Collocation errors are still captured, but under the more inclusive [Phrase] category.

6.5.3 The next steps

The next chapter, Study 5, focuses on testing the inter-rater agreement of NewLEAF2, but on a larger number of participants with a greater variety of qualifications and experience to produce NewLEAF3. The following chapter (Study 6) will seek to compare NewLEAF3 with the most recent comparable framework: that created by Hemchua and Schmitt (2006) to ascertain whether the new framework is felt to be an improvement in any way to its predecessor. This was done to some extent with NewLEAF1 in Study 2. The comparison will be done by using the same 20 essays produced by Greek learners that were used in Study 1: the replication.

Chapter 7 Study 5: Testing and Refining NewLEAF2

7.1 Introduction

Study 4 sought to qualitatively trial the new framework and gather thoughts on it from practising English language teachers. It was found that while these teachers stated that they thought the framework would have various uses in their practice, and they were confident in their categorisations, the inter-rater agreement scores, i.e., the similarity of results between participants when identifying and categorising lexical errors, were too low and required improvement. This led to the redesign of the framework and guidance in line with their comments and instances where there was dual categorisation of LEs: the result is NewLEAF2.

In order to test NewLEAF2, this study starts with a repeat of Study 4, with the same participants and same student essay with errors, but seeks to collect quantitative data only. It was conducted to measure whether there are any improvements in agreement scores following the design of NewLEAF2. This repeat of aspects of Study 4 is hereafter referred to as Part One. Part Two is the same quantitative study again, but this time with a wider variety of Applied Linguists, such as linguistics lecturers, more TESOL tutors, students of TESOL and Linguistics, including participants who have English as a second language. These investigations into the inter-rater agreement can be justified by the fact that, to date, there have been no inter-rater agreement studies in LEA. Where various authors, such as Dušková (1969), Richards (1971), Corder (1973), Zimmerman (1986), Meara and English (1987), Zimmerman (1987), Lennon (1991), Zughoul (1991), Engber (1995), James (1998), Hemchua and Schmitt (2006), and Llach (2011) have produced original frameworks for LEA, they have not been tested with different raters. It is important for the framework to have high inter-rater agreement, to ensure it can be used confidently by teachers and researchers to compare findings of LEA across different groups or individuals, or by multiple teachers and researchers to measure changes over time with the same group or individuals.

In the sections that follow, I present the research questions, explain the methodology and compare and discuss results from the different sub-groups. These results and inter-rater agreement issues are then analysed for opportunities to improve the framework and guidance to produce NewLEAF3.

7.2 Methodology

This section will present the research questions, justify the methodology, describe the participants and explain the data collection and analysis techniques. It is divided into Parts One and Two.

Part One, a quantitative only repeat of the previous study was conducted to measure any increases or decreases in inter-rater agreement with the new framework, using the same variables. It was important to conduct a replication so that any improvement in agreement scores could be measured accurately by controlling as many variables as possible. However, there were some minor amendments to address perceived methodological weaknesses in the previous study. These are described below (see Section 7.2.2.2).

Further trialling is investigated in the current study. Part Two comprises a wider investigation: again, quantitative data only were collected from a larger number and wider range of participants. These participants and ethical procedures are described below. This is followed by explanations of data collection and analysis procedures. In Part Two, 38 participants identified errors in a piece of writing and 33 of those categorised the extracted LEs using NewLEAF2. Participants were divided into two subgroups. Group A (18) comprised qualified, experienced, L1-speaker English language teachers, and Group B (15) was made up of various other participants from within the Applied Linguistics field, including qualified but inexperienced language teachers and those who had another first language, Master's degree students (TESOL and Linguistics) and lecturers in linguistics.

7.2.1 Research questions

- 1) How similar are the results when participants are asked to identify LEs in a piece of English language learner writing when using the new guidance (NewLEAF2)?
- 2) How similar are the results when participants are asked to categorise a series of extracted LEs using the new guidance and framework (NewLEAF2)?

Research aim

If there is much difference found between participants when answering RQs 1 and 2 above, how can the framework be further refined?

7.2.2 Part One

7.2.2.1 Sampling, participants and ethics

Convenience sampling, as discussed in Section 6.2.3.1 was used for Part One. The same participants that were used for the current study were used in Study 4 (see Section 6.2.3.3 for details). It was important to control the variable of intra-rater agreement. If different participants from those used in the previous study were used, it could have introduced issues such as varying ability to follow instructions and differing beliefs about what constitutes LEs. The participant information sheets and consent forms that the participants above signed for the previous study covered ethical requirements for the current study also (See Appendices 7.1 and 7.2).

7.2.2.2 Data collection

As mentioned above, Part One is a replication of Study 4. The same participants were asked to identify LEs from the same essay and then categorise the same 37 previously extracted errors. The six-month gap between the data collection for Study 4 and the current study was sufficient in that participants would be unable to recollect their responses when identifying and categorising LEs. Despite some concerns in Study 4 that asking participants to discuss identification and categorisation in pairs would affect their initial decisions in these areas, it was decided to continue with this approach, as voicing their thoughts would replicate Study 4 and help them understand each other and follow instructions.

There were only three slight differences in the data collection methods for this replication: 1) In Study 4 it was important that participants receive written instructions for task completion to speed up data collection on the day. It was also felt that if LEA was to be conducted beyond the environment of the trial, written instructions had to be issued as I could not always be there to explain the procedures. However, during the last study, it became apparent that not all participants had read or fully understood the emailed guidance and framework

before they were asked to complete the tasks. Therefore, in this instance, I pre-recorded two screencasts that carefully explained firstly how to identify LEs and secondly how to categorise the extracted LEs using the new framework (NewLEAF2). Participants were played these videos before they were asked to complete the two tasks of identification and categorisation. This also ensured that all participants received exactly the same instructions: an important variable to control to ascertain a more accurate inter-rater agreement score. The two videos can be viewed here:

- 1) [Task 1, Lexical Error Identification](#)
- 2) [Task 2, Lexical Error Categorisation](#)

2) Written instructions (printouts of Appendix 7.3) were also provided for reference during the analysis so that participants did not have to rely on their memories to complete the tasks in line with the guidance. 3) Participants were also provided with a printed copy of the original essay from which the errors were extracted so that they could consult it should they struggle to comprehend the meaning behind some of the extracted errors. It was thought that the added context may help clarify the meaning of some of the errors and create more of a real-world situation, as analysts using the framework would probably be using complete texts, not extracted errors. These were the only differences in data collection between Study 4 and this study. It was decided not to conduct post-task interviews, as sufficient data to answer the interview questions had already been obtained. Instead, further quantitative data was sought in Part Two (see Section 7.2.3).

Identification

In the first task, again, participants were asked to think aloud as they identified LEs. It was hoped that this would replicate Study 4 and their thoughts would guide each other to arrive at considered identifications. As previously mentioned, they were supplied with the instructions and guidance in Appendix 7.3, which included the same learner essay used in the previous study, and they were asked to underline the LEs in an MS Word document on screen.

Categorisation

For the second task, they were asked to categorise the same 37 extracted lexical

essays from the same essay, again using MS Word to record their categorisations and confidence scores (See Appendix 7.3 for instructions, framework and guidance). The participants would not have been able to recall and use the same categorisations as the framework was different and six months had passed since the data collection for Study 4. Again, to replicate Study 4, and to help each other arrive at considered categorisations, participants were asked to discuss the categorisations before recording individual decisions. They were also provided with a printed copy of the framework for reference during completion of Task Two.

7.2.2.3 Data analysis - Error identification, error categorisation and confidence scores

The same procedures for data analysis that were used in Study 4 were used here (See Section 6.2.5). The percentage agreement figures were compared to the corresponding figures from the previous study to establish whether any improvements in agreement scores had been achieved with NewLEAF2 in these areas.

7.2.3 Part 2

Part Two incorporates a greater number of participants and repeats the procedures above. It did not seek to collect qualitative data for the reasons for the participants' identifications and categorisations, nor on their opinions on how the framework and guidance could be used or improved. Quantitative data of their actual decisions were sought to establish the inter-rater agreement of the identification and categorisation aspects of the framework.

7.2.3.1 Sampling

It was intended that convenience sampling (See Section 6.2.3.1) could be used in conjunction with snowball sampling for this study. Participants known to myself were invited to join by email, and they were asked to forward the email to their industry contacts in the hope that they would participate in the study also. Unfortunately, no further participants were recruited in this way.

7.2.3.2 Participants

Participants were allocated to two broad groups, as shown in Table 40 below.

Code	Types of participants	Number of Participants for Task 1 /Task 2
Group A		
EFLQEL T	English first language, qualified, experienced English Language teacher	21/18
Group B		
Practising English Language Teachers		
EFLTTL	English first language, trainee English Language teacher	1/1
ENFLQE LT	English not first language, qualified, experienced English Language teacher	11/9
Practising Linguists (not working as English Language Teachers)		
ENFLQL T	English not first language, qualified English linguistics teacher	1/1
EFLQLT	English first language, qualified linguist	3/3
EFLTTL	English first language, trainee linguist	1/1
Group B Total		17/15

Table 39 Participants in Study 5

The following subgroups were originally created to reveal any differences in agreement when the framework is used by different subgroups, not to establish any perceived better performance by one group over another. As there was an unequal response to the call for participants from some subgroups, some were merged. Subgroup A was created as the call for participants had attracted a sizeable number to populate an individual group. Subgroup B was created by merging the other participants together to create a similarly sized group. Unfortunately, the call for participants did not attract large enough numbers to enable sizeable groupings of further subgroups. Further, English L1 speakers and non-English L1speakers were merged together. There are differing numbers of participants for the two tasks as not all participants returned their responses to Task Two.

Subgroup A - English first language, qualified, experienced, practising English language teachers

The main sub-group of participants had English as a first language, held at least an initial qualification in English language teaching and also had English language teaching experience. This was the largest group for convenience sampling

reasons: they were the largest sub-group of participants known to me as contacts. They, along with those teachers who did not have English as a first language, would probably be the main group of future target users for the framework.

Subgroup B - English first language, trainee English language teachers

Trainee English Language teachers without a qualification or experience were also initially asked to participate to see if a lack of qualification or experience led to any differences between them and those that had experience and a qualification.

English not as a first language teacher, qualified, experienced English language teachers

This group, who did not have English as a first language but were qualified and experienced English Language teachers were originally included to test the framework on highly skilled language users that did not have English as a first language.

English first language, qualified linguist

This group of participants comprised those who had English as a first language and held a qualification in linguistics, but were not language teachers. This subgroup included university lecturers. They were included as one application of the framework would be for linguists to use to analyse language use within or between groups of learners.

English not first language, qualified English linguistics teacher

This subgroup, those linguistics teachers who held a qualification in linguistics, but did not have English as a first language, was also initially invited to act as a comparison group.

English first language, trainee linguist

This subgroup comprised of those who had English as a first language and were studying for a Master's degree in linguistics. They were initially invited to see if linguistically aware non-teachers could use the framework in the same way as experienced, qualified English Language teachers.

7.2.3.3 Ethics

Full ethical approval was granted by the Ethics Committee at Manchester Metropolitan University (see Appendix 7.4 for approval letter). All participants in Part Two were asked to read the participant information sheet (See Appendix 7.1) and sign the consent form (See Appendix 7.2).

7.2.3.4 Data collection

A questionnaire (See Appendix 7.5) was designed to collect data on first language, qualifications in English language teaching or linguistics and number of years' teaching experience. All participants could be described as highly-skilled users of English, as they were either employed as lecturers, tutors or were enrolled in higher education courses. They were asked about their occupation (student, tutor or lecturer), qualifications and number of years' experience to categorise participants into the groups above. Data collection for the two tasks, described in Section 7.2.2.2, was done in two phases: face-to face and by email.

Face to face

50 participants from the groups above were invited by email to a large classroom. 15 came to the data collection session. They were allocated a participant number on arrival and were asked to record it on all documents to be returned to me. This enabled participant anonymity and me to track which documents were completed by whom. They were asked to sit apart from each other so that they could not see each other's responses, to reduce the likelihood of copying or sharing answers. They were asked to complete the tasks immediately after each of the two videos had been played. They were given 10 minutes for the first task and 20 minutes for the second. Their completed task sheets were collected after each task. This was done in two phases as I did not want participants to amend their first task responses (identification of all LEs in a student essay) after they had seen the extracted errors in the second task.

Email

Those who could not attend the face-to-face data collection session due to other commitments, but who had expressed an interest in taking part in the study, were emailed the participant information sheet, consent form and questionnaire and

Task One document (See Appendix 7.6) and also the link to the first instructional video. They were also asked to forward the email onto any contacts they may have who would be suitably qualified and willing to participate. After they had returned the completed documents to me, they were emailed the second set of instructions, video link and categorisation task document to complete and return (See Appendix 7.7). A higher response rate could be achieved if the two tasks were separated, making them appear less onerous. Participant numbers were allocated on receipt of their first email.

7.2.3.5 Data analysis

Identification

The same procedure used in Study 4 was replicated here (see Section 6.2.5). However, in this phase, participant responses were grouped according to the participant subgroups (see Section 7.2.3.2) and differences between these groups were analysed. Firstly, to form a larger group, Subgroup A (the EFLQELTs) responses were added to the corresponding data set from Part one and the data were analysed to establish an overall inter-rater agreement score for this subgroup. The same was done for Subgroup B. Next, the average inter-rater agreement score for all participants for identification was calculated.

Categorisation

Again, the same procedure that was used in Study 4, was replicated here (see Section 6.2.5), and again participant responses were grouped according to Subgroups A and B, and differences between these groups were analysed. Firstly, Subgroup A responses were added to the corresponding data set from Part One and the data were analysed in the same way to establish an overall inter-rater agreement score for this subgroup. The same was done for Subgroup B. Next, the average inter-rater agreement score for all participants for categorisation was calculated.

Confidence scores

Again, the same procedure that was used in Study 4, was replicated here (see Section 6.2.5), and again participant responses were grouped according to

Subgroups A and B, and differences between these groups were analysed by comparing the two to see which was the greater. First, Subgroup A responses were added to the corresponding data set from Part One and the data were analysed in the same way to establish an overall inter-rater agreement score for this sub-group. The same was done for Subgroup B. Next, the average inter-rater agreement score for all participants for categorisation was calculated.

Research aim

If there is much difference in responses to research questions 1 and 2 above, how can the framework be further refined?

Identification

To better understand how to promote better agreement in error identification between participants, errors with low agreement rates were analysed for factors that led to their identification by only a few participants. These factors are then listed in Section 7.3.2.1. and then addressed to modify the guidance to exclude inconsistencies.

Categorisation

To establish how the categorisation aspect of the framework could be improved, it was decided to analyse Dual Categorisation Instances (DCIs). As Group A provided a set of results with a higher agreement figure than Group B (several of the categorisations from Group B seemed quite arbitrary), the data set from Group A categorisations was selected to analyse DCIs. 30 out of the 37 extracted errors (112 DCIs) from Part Two were analysed. To be included in this part of the analysis, these errors had to have an agreement score of 72% or lower or had two or more 'next most common' modes. These two figures were selected as they represented a certain cut off point of acceptability. Not all DCIs were investigated: a perfect system of categorisation could not be attained and if adjustments were made to aspects of the framework that were working relatively well, then those adjustments may cause issues in areas that were functioning well. Where there were fewer than two disagreements with the mode, these were considered outliers and excluded from the analysis. An error code was allocated to each of these DCIs. The number of these codes were added up to find the most common issues

with NewLEAF2. It was hoped that suggested amendments to be made to the framework may eradicate or reduce these issues.

7.3 Results and discussion

This section is organised by Part One and Part Two of the study and then by research questions within those parts.

7.3.1 Part 1

7.3.1.1 Identification research question 1 - How similar are the results when participants are asked to identify LEs in a piece of English Language learner writing when using the new framework and guidance (NewLEAF2)?

As can be seen from Table 41 below, the overall inter-rater agreement score for the current study was 51.36%, whereas in Study 4, it was 59.3%. This represents a decrease of 7.94%: it was hoped that an increase following improvements to NewLEAF1 would be established. However, it must be stated that with such low numbers of participants, these agreement scores could not be considered important. This decrease can be explained mainly by the fact that the participants identified 12 more errors in the current study than they did in the last, and as not all participants agreed that they were errors: it reduced the inter-rater agreement score considerably. For example, of the 55 errors that were identified by all participants, there was 100% agreement for only seven errors: Nos 1 and 28 [Verb misselection], Nos 9, 27, 33 and 37 [Phrase misselection] and No 15 [Preposition misselection], 75% agreement for nine further errors, 50% for 19 errors and 25% agreement for 20 of them. (See Table 7.3 in Appendix 7.8). It is doubtful that the reduction in agreement could be explained by the slight differences in methodology: participants were given videoed and written instructions in the current study in how to identify and exclude grammatical errors from the analysis, rather than written only instructions in Study 4. Interestingly, the number of errors identified by two pairs of participants increased in the current study (Participants 5 and 6 identified almost twice as many errors) and the number identified by one pair decreased. Interestingly, of the six extra errors identified by Participants 1-6 in Study 4, none were repeated in Study 5: they identified completely different errors. This suggests that memory did not play a role in Study 5, but raises questions about intra-rater reliability in the identification of LEs.

Participant	Number of LEs (V1: Study 4)	Number of LEs (V2: Study 5)
Researcher	37	37
1 and 2	20	25
3 and 4	30	24
5 and 6	15	27
Mean	25.5	28.25
Tot no different LEs	43	55
Agreement score	59.3%	51.36%

Table 40 Comparison of error identification with NewLEAF1

Possible further issues affecting agreement of identification

As in Study 4 (see Section 6.3.1), participants exhibited various levels of strictness when identifying errors. I identified 37 errors, 10 more than the next highest identified number (by Participants 5 and 6). Perhaps the participants had been schooled in the principle of not correcting all errors in a piece of student work, as it may lead to demotivation (Harmer 2015), and this possibly reduced the number of errors that they felt comfortable in identifying in comparison to the amount I identified. This difference in strictness also added to the low agreement score. Consistency appeared to be an issue: even I was unable to maintain consistency in strict formality error identification, for example. I failed to identify Error 54 '*Above all, I think, the most important.*' (see Table 7.3 in Appendix 7.8). Ability to pay attention to detail may also be an issue, as Participants 5 and 6, for example, failed to identify '*performan*', and '*undebateably*' as errors.

7.3.1.2 Categorisation Research question 2 - How similar are the results when participants are asked to categorise a series of extracted LEs using the new guidance and framework (NewLEAF2)?

The overall inter-rater agreement score for NewLEAF2 was 80.11% (See Appendix 7.9 Table 7.4 Part One Error Categorisation and Confidence Scores), which could be seen as a satisfactory score. For NewLEAF1, it was 66.54%. This represents an increase of 13.57%. This improvement may be due to the

eradication of some dual-categorisation possibilities: chiefly by removing the highly-subjective options to classify errors as [Wrong word (field)] error or [Wrong word (sense relations)] error, and also the clarification of how to categorise errors with single words or phrase errors, as well as by addressing other issues. Interestingly, the mode of the categorisations, found by comparing the categorisations of all seven participants, matched the code allocated by myself (see Appendix 7.10, Table 7.5). This shows that the framework has some degree of inter-rater reliability, and it is largely being used for categorisation as intended, but this will be further investigated when the same test is applied to data from a wider number of participants in Part Two.

15 errors (41% of all errors) were found to have 100% categorisation agreement between all seven participants, 18 (49%) had an agreement score of 71%, one error (3%) had an agreement score of 57% and three errors (9% of all errors) had an agreement score of 43%. Overall, 33 errors returned high agreement scores (71% and over) and four errors returned low agreement scores (51% and below). These results are very encouraging and support LEA as a useful tool for teachers. However, this high score could, again, be due to the fact that pairs could discuss categorisations before agreement. Also, it must be acknowledged that this is the second time that these participants from a homogenous group had been asked to perform LEA, so perhaps higher agreement could be expected.

The Dual Categorisation Issues (DCIs) that led to the low agreement scores for individual errors will be investigated in Section 7.3.2.4, along with the corresponding data from Part Two.

Confidence scores

The totalled average confidence score for all participants and the totalled average confidence score for all errors are the same within a study as they are calculated using the same figures. However, the average confidence score for the current study, using NewLEAF2, was 2.77 (See Appendix 7.9 Table 7.4 Part One Error Categorisation and Confidence Scores), which demonstrates that the participants felt confident, overall, with their categorisations. The average confidence score for NewLEAF1 was 2.68. This represents an increase of 0.09. Whilst this is a step in

the right direction, it is gained from a small number of participants. Perhaps the same factors that led to a greater similarity in agreement for the categorisation task were also responsible for this slight rise in confidence scores. However, it is not a large increase and does not therefore warrant much speculation as to why the increase occurred. Interestingly, across both studies, where there is 100% agreement on the categorisation of individual errors, the lowest average confidence score is 2.71. Where there is the lowest agreement (29%), confidence scores do not rise above 2.71. This indicates that there is some correlation between agreement and confidence.

7.1.3.3 Research aim - If there is much difference in 1 and 2 above, how can the framework be further refined? This will be addressed in Section 7.3.2.3.

7.3.2 Part Two

Part Two of this study incorporates a wider range of participants, described in Section 7.2.3.2.

7.3.2.1 Identification research question 1 - How similar are the results when participants are asked to identify LEs in a piece of English Language learner writing when using the new framework and guidance (NewLEAF2) for identification?

Results

Across the two groups, A and B, 154 errors were identified. The overall inter-rater agreement score was very low (18.69%). It was interesting to note that each participant identified errors that others did not. Individually, there was even greater variety in the total number of errors identified by the participants, which was perhaps to be expected when using a larger number of participants. In Study 4, a total number of 43 errors were identified by all participants. In the current study, it was 55 in Part One and, as mentioned, 154 in Part Two.

Individually, the total number of errors identified varied remarkably between participants: from 8-68. (See Appendix 7.11 Table 7.6 Part Two Group A and Group B Error Identification.) The mean number of errors found was 29 (the

slightly larger Group A identified 665 errors in total, whereas Group B identified 440. The Group A average was 32 per participant, and Group B's was 26).

The overall inter-rater agreement figures improve slightly when analysing the two groups individually. Group A identified a combined total of 665 errors (136 different errors) and showed an agreement score of 23.28% (see Appendix 7.12 Table 7.7 Part Two Group A Identification). Group B identified a combined total of 440 errors (121 different errors) and saw an agreement score of 21.39% (see Appendix 7.13 Table 7.8 Part Two Group B Identification). Low error agreement scores appear to be affected by several factors. These are discussed below with examples:

a) Raters' beliefs about what constitutes an error

In Group A, 23.8% of participants identified '*...compared *with others...*' (Error 38) as an error. This appears to be a matter of preference, not error. Perhaps to the majority, '*...compared to others...*' is equally suitable.

b) Differences between raters' acceptance of different varieties of English

Error 40 '*how to read them *fast*' (American English) was identified by 33.3% of participants.

c) Raters' beliefs about social acceptability

Error 50 (*for every *foreign student*) was deemed as such due to the politically incorrect use of the word 'foreign'. 19% saw this as an error.

d) Raters' acceptance of tautological expression

28.6% of participants identified Error 90 '*..is a *very crucial..*'.

e) Raters' beliefs about what constitutes an error in one genre but not another

Only 47.6% identified Error 1 (*To *get a perfect academic performan is a basic quality that every student wants*) as an error. The issue here is varying acceptance of degrees of formality within a genre.

f) Raters' ability to concentrate and identify errors consistently

Even I failed to consistently apply my own understanding of acceptable formality in academic writing as I missed Error 54 '*Above all, *I think, the most important...*' 14.3% felt this to be an error.

g) Raters' ability to follow instructions in the guidance

Participants were asked to exclude errors with plurals, yet 14.3% identified Error 70 as such '*the listening skill is a prior skill for a *students*'.

h) Raters' subjective beliefs about the blurred boundary between lexis and grammar

23.8% of participants identified Error 64 '*Not only *reading skill can help student doing*'. This appears to be a syntax, or word order error, which some would have felt to have been grammatical and therefore excluded, but it could also have been seen by some participants as a misordering error of a relatively common fixed lexical phrase (i.e., '*Not only can xxx, but it can also yyy*'). Also see Error 65. These are probably weaknesses in the guidance when participants try to separate grammatical and lexical errors.

i) Differences between raters' identification of the extent of the error

It is not recorded in Appendix 7.12 Table 7.7 Part Two Group A Identification, but it was seen as the data were compiled that different parts of a phrase or the whole phrase could be identified as an error. Participants have identified Error 1 ('*To *get a perfect academic *performan* ') differently, as an [informality], [spelling] and a [phrase] error.

Corresponding suggested actions to issues above:

- a) Include the following in the guidance, "Include errors of variety, e.g., American English"
- b) Include the following in the guidance, "Include errors of social awareness, e.g., Foreign Student".
- c) Include the following in the guidance, "Include errors of tautological expression, e.g., Very crucial"
- d) Highlight the instruction to identify errors of genre.
- e) Request that analysts scan the text twice for errors.
- f) Reduce word count in the guidance to help analysts understand and retain advice and information. Ask them to consult the guidance as they analyse.
- g) Ask analysts to include word order errors in fixed expressions (phrases) only.
- h) Ask analysts to underline the extent of the erroneous writing within a phrase. However, this may increase agreement scores for identification, but

in the categorisation phase, may mask the actual types of errors being made and artificially swell the number of phrase errors in the analyses. See also Errors 40, 93, 140, 141, 144, 150 below for further examples of issues with extent of error identification.

Discussion

These findings point mainly to weaknesses in the guidance and its inability to ensure raters identify errors in the same way. The definition currently used of what constitutes an LE uses more words on what an LE is not, rather than what it is. Integrating the suggestions above may increase agreement.

The descriptive statistics in the section above tell us that Group A identified more errors than Group B overall and per participant. It is not clear why this difference exists, but one reason could be that Group A, which comprised experienced, qualified, L1-speaking English Language teachers, was a more homogeneous group and therefore identified in a more similar way. The design of Study 4 and Study 5 did not seek to collect qualitative data on the reasons why raters identified LEs differently from each other.

It is of limited value to compare the results of the current study to those from previous studies of error identification, as these may have focussed on identification of grammatical errors. Grammatical accuracy is more governed by prescriptive rules, whereas lexis appears to be less so. Therefore, one might expect lower agreement scores when analysing LE identification agreement between participants. It would seem impossible to produce a reference manual that would allow for such norming of identification, as it would involve a vast and unwieldy range of all words and phrases with quite subjective judgements on correctness in different genres. Further, my recommendation from Study 4 called for less information in the guidance and framework to simplify the process of LEA. However, the guidance could usefully be amended to include the following to improve consistency in identification: 'include words or phrases that you consider to be awkward expression, as well as words or phrases that you consider to be erroneous in that genre'. This represents a brief addition to the guidance, but implementation may still vary and may even cause confusion when considering formality errors and overspecification errors in academic writing, for example. The low agreement score of 18.69% for identification does not support LEA studies

where more than one analyst is used. Intra-rater agreement is not investigated within this dissertation and there do not seem to be any published intra-rater LE identification studies. This is a direction for possible future research.

Fortunately, based on the findings from the current study, there are opportunities to improve the framework for categorisation purposes.

7.3.2.2 Categorisation research question 2 - How similar are the results when participants are asked to categorise a series of extracted LEs using the new guidance and NewLEAF2?

Results

Again, across both groups, the combined overall agreement scores for both groups' categorisations were very low (55.32%), which was quite disappointing, given the high agreement score from the corresponding study in Part One (80.11%). Group A's score was 64.11% and Group B's was 48.83% (see Appendix 7.14 Table 7.9 Part Two Group A and Group B Error Categorisations with Mode). The mode for Group A matched mine, as did the mode from the corresponding study in Part One in 34 out of 37 instances (Errors 12, 30 and 34 did not). The modes from Group B matched the modes in Group A in 30 out of 37 instances (81%). Whilst the inter-rater agreement figures for Groups A and B can be considered low, the similarity in modes overall, points to satisfactory agreement levels.

Discussion

It seems that agreement percentages drop with a larger number of participants and that Group A can allocate more consistently than Group B. Again, it is difficult to speculate as to the reason behind this difference. It could be due to the fact that Group A have similar backgrounds in terms of their duties in line with their work and are more likely to categorise in similar ways. Dual-categorisation issues (DCIs) are dealt with in more detail in Section 7.3.2.4.

A subsidiary hypothesis was that agreement scores would increase during the analysis as participants get used to using the framework. However, this did not

seem to be the case. It seems agreement scores were more affected by type of error: the lower the agreement, the lower the confidence scores.

Confidence scores

Confidence scores also varied between the groups. The overall score out of three was 2.38: for Group A it was 2.41, and for Group B it was 2.35 (See Appendix 7.15 Table 7.10 Part Two Group A and B Categorisation Confidence Scores). This points to the fact that despite the low inter-rater agreement scores for categorisation, individual participants felt quite confident in their categorisations. Perhaps Group B overall felt less confident because the majority of them were working in a second language. As groups' confidence scores seemed to increase in line with agreement scores for individual errors, improvements in confidence scores could probably be increased with improvements to the framework to clarify the guidance and reduce the possibility of dual categorisation (see the next section).

A subsidiary hypothesis was that confidence scores would increase during the analysis as participants get used to the framework. However, as above, this did not seem to be the case. Confidence scores were more affected by type of error; the higher the agreement, the higher the confidence scores. For example, spelling errors had the highest agreement scores and confidence scores (errors 2, 7, 13, 25 and 26). Errors 30 and 34 both had some of the lowest scores in both areas.

To establish whether there is a correlation between average confidence score per error and percentage agreement of categorisation in line with the mode, both sets of data were ranked and a Spearman's correlation (non-parametric) analysis was run. This returned an output of 0.73, which indicates a strong, positive correlation between the two variables in line with the hypothesis. The association between the variables is statistically significant also. See Tables 42 and 43 below.

Correlations				
		Avg_Confidence		Perc_Agreemen t
Spearman's rho	Avg_Confidence	Correlation Coefficient	1.000	.731**
		Sig. (1-tailed)	.	<.001
		N	37	37
	Perc_Agreement	Correlation Coefficient	.731**	1.000
		Sig. (1-tailed)	<.001	.
		N	37	37

** . Correlation is significant at the 0.01 level (1-tailed).

Table 42 Spearman's rho correlations between confidence and agreement scores

Strength of Association	Coefficient, <i>r</i>	
	Positive	Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.3 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to -1.0

Table 43 Statistical significance coefficient

7.3.2.3 Improvements to the framework research aim - If there is much difference in 1 and 2 above, how can the framework be further refined?

This section discusses what could be done to improve inter-rater agreement for LE identification and categorisation, and what associated refinements could be made to the framework.

Identification

Due to the issues raised in Section 7.3.2.1, identification appears to be the most problematic aspect of LEA. Whilst a manual that offers guidance on the 'correctness' of lexis would be subjective, contentious and unwieldy, as previously stated, perhaps expanding the guidance to the following sentence: 'Only include words or phrases that are completely wrong for the genre. Do not include words or phrases that you consider to be merely not well chosen.' However, this would not completely eradicate subjectivity (what may seem odd to some, may not be odd to others). Indeed, the term 'highly-skilled' in the definition of a LE is also subjective. Instead, it would be better to amend the definition to read: 'A written lexical error is

a lexical form or combination of forms, which would not be made by a writer who has attained IELTS Band 9 (the highest score), writing in the same genre'. Although it is doubtful that this would completely eradicate disagreement, the best one could hope for is good levels of agreement, perhaps 80%. It is probably impossible to completely avoid subjective judgements. If it was not, then automated systems, such as Microsoft Word or Grammarly would be 100% flawless, which they are not.

In terms of extent of error identification, some participants underlined an erroneous word in a phrase, others underlined more than one word when identifying the same error. Inter-rater agreement could be improved if this issue were also addressed in the guidance. Suggested action: include the following in the guidance: Underline all the LEs in the writing: both single word errors (*I was very *happiness to receive the present.*) and only the erroneous parts of whole phrases (*It was raining *dogs and cats*). This would be one solution to the issue of identification of extent of error (see Lennon 1991).

Categorisation

To enable analysis, the instances of Dual Categorisation Issues (DCIs) below were mainly grouped by mode. For example, all DCIs that occurred when [Verb misselection] was the most common categorisation choice were grouped to establish common issues. Out of 112 instances of dual categorisation that were selected for investigation, the highest number (36) related to errors that most participants had categorised as [Formality] errors. The second largest area was errors relating to [Phrase misselection] (24), then [Coherence] and [Noun] (14 each) then [Spelling] (12), then [Verb misselection] (10), and finally [Two words should be one] (2). See Table 44 below.

Description of Dual Categorisation Issue	Code	Fre- quency
Formality error and Phrase misselection error	DFPM	13
Formality error and error of Pronoun omission	DFP	3
Formality error and Uncategorisable error	DFUN	13
Formality error and error of Underspecification.	DFU	4
Single word (misc) selection error and error of Formality	DSF	3
Total No instances of dual categorisation including Formality		36

Description of Dual Categorisation Issue	Code	Fre- quency
Phrase misselection error and error of Adjective misselection	DPA	4
Phrase misselection error and error of Coherence	DPC	5
Phrase misselection error and error of Verbosity	DPVE	2
Phrase misselection error and error of Verb misselection	DPV	3
Phrase misselection and error of Underspecification	DUP	10
Total No instances of dual categorisation including Phrase misselection		24
Coherence error and error of Phrase (misc) disorder	DCPM	3
Coherence error and error of Underspecification.	DCU	8
Coherence error and error of Suffix omission	DCS	3
Total No instances of dual categorisation including Coherence		14
Noun overinclusion error and error of Two words should be one (blend)	DNB	2
Noun misselection error and error of Single word (miscellaneous) misselection	DNM	8
Noun overinclusion error and Uncategorisable error	DNU	2
Noun overinclusion error and Noun misselection error	DNN	2
Total No instances of dual categorisation including noun overinclusion		14
Spelling misselection error and Spelling blend error	DSSB	2
Spelling misselection error and error of Spelling overinclusion.	DSSO	3
Spelling error and error of Suffix omission	DSS	3
Spelling misselection error and error of Suffix misselection	DSSM	3
Total No instances of dual categorisation including spelling		12
Verb misselection error and Preposition omission error.	DVP	2
Verb misselection error and error of Single word (miscellaneous) misselection	DVS	8
Total No instances of dual categorisation including Verb misselection		10
Two words should be one error and Uncategorisable	DTU	2
Total No instances of dual categorisation including Two words should be one		2
Total No instanced of dual categorisation investigated		112

Table 44 Dual categorisation issues, codes and frequencies

The next sections will analyse these dual categorisations by specific error type and dual categorisation code to establish whether improvements could be made to NewLEAF2.

Formality

DFPM [Formality] error and [Phrase misselection] error

Error 15. *'I must use lots of knowledge and documents *that I do not know to make it better.'*

Error 16. *'I must use lots of knowledge and documents that I do not know to *make it better.'*

Error 37. *'the best way to *obtain the biggest progress'*

A high number of participants (13 out of the 41) dual categorised these errors differently from the mode. It is quite understandable that there is dual categorisation here as different participants could have had different concepts of formality: perhaps they could all be seen as both [Formality] and [Phrase misselection]. The framework should state that [Phrase] errors that relate to [Formality] should be categorised to the latter, as it is the more specific. Suggested action: make [Formality] a subcategory of [Phrase] error and encourage participants to allocate to the more specific category: [Formality].

DFP [Formality] error and error of [Pronoun omission]

Error 4. *'Let *me discuss the topic'* It is unclear why three participants saw this [Formality] error as a [Pronoun omission] error. Perhaps they did not see it as an error (informal use of reference to self). Perhaps it was categorised erroneously in haste. Certainly, the 'me' is the problematic aspect, but the error cannot be fixed by replacing this one word alone. No suggested action.

DFUN [Formality] error and [Uncategorisable] error

Error 19. *'So, in *my opinion'*

Error 23. *'Before *you use the new knowledge, you need hear the content of what proffesser says clearly.'*

Error 17. *'*Is it the most important ? I do not think so.'*

Error 16. *'Is it the most important ? *I do not think so.'*

An alternative way of grouping these data would be to examine the large number of uncategorisable errors. One solution to reduce this number would be to remove the [Uncategorisable] category altogether, but this would hide the issue, rather than fix it. Removing the category would address 17 of the DCIs in Table 42 by forcing analysts to think again and perhaps categorise in line with the mode. However, this would remove the option of having a place for analysts to place errors that they are struggling with or any that cannot be categorised despite the fact that they may indeed be categorisable in line with the guidance. Retaining the category may serve as a place for analysts to allocate to if they have not fully understood or read the guidance. It was important to retain the category as it acts as an ‘error bin’ to collect possible future miscellaneous errors as well as provide evidence of misunderstanding or lack of implementation of the guidance. This information will enable future improvements to the framework and ensure all errors that have been identified are included in the analysis. This issue is somewhat similar to the debate with Likert scales – do you include a middle ‘not sure’ answer, or force people to make a choice? The ‘error bin’ was decided upon because it tells the learner that there is a problem, even if it cannot confidently be categorised, so at the very least it needs rephrasing.

The largest number of issues relating to these errors relate to [Formality] also. This number was increased by the inclusion of the errors above that were not seen as errors by the majority. Suggested action: reduce the word count in the guidance and framework and highlight request to identify errors, based on genre.

DFU [Formality] error and error of [Underspecification].

Error 17. *“Is it the most important ? I do not think so.”* Perhaps these five participants, who saw this [Formality] error as an error of [Underspecification] were unaware of the fact that rhetorical questions are widely frowned upon in academic writing (University of Portland, 2023) or were unaware of what ‘it’ (Reading Skills in the error above) referred to. They were asked to consult the original co-text, which would have made the meaning clear. I have no further suggestions to make the meaning of the extracted errors clearer.

DSF [Single word (misc) selection] error and error of [Formality]

Error 30. *“Whatever a student hear and write and how fast a student read, if a*

student can not talk own views with others, it is vein for the student to have so many good skills.' Four participants considered this [Single word (miscellaneous) selection] error an error of [Formality]. Perhaps they considered 'Whatever' to be a stylistic error more common in speech. However, to avoid future DCIs between these two categories, suggested actions would be to remove the redundant [Single word (misc)] category and ask analysts to allocate [Single word formality] errors to the [Formality] category, as this is the more specific description of the error.

Errors with phrases

DUP [Phrase misselection] and error of [Underspecification]

Error 11. '*I must use lots of knowledge and *documents that I do not know to make it better.*'

Error 20. '*So, in my opinion , *it is the most direct approach to gain great academic performance.*'

Error 27. '*No tutor teaches students only by writing on blackboard without *any voice.*'

10 participants considered these errors of [Phrase misselection] errors of [Underspecification]. Perhaps they reformulated the error to use more words than the student did. Suggested action: Make [Underspecification] a sub-category of [Phrase] error and encourage analysts to allocate to the more specific category: [Underspecification].

DPV [Phrase misselection] error and error of [Verb misselection]

Error 21. '**to gain great academic performance.*' Three participants considered this [Phrase misselection] error an error of [Verb misselection]. This categorisation ignored the fact that the whole phrase was highlighted as an error. These participants focussed on only one word. This is another instance of participants not following instructions, which was probably due to cognitive overload. Suggested action: reduce word count in guidance and framework and simplify the process.

DPA [Phrase misselection] error and error of [Adjective misselection]

Error 22. '*, the listening skill is a *prior skill* for a students.' Four participants considered this [Phrase misselection] error an error of [Adjective misselection]. There certainly is an erroneous adjective, but the whole phrase was in italics and

to fix the error, it would require more than simply replacing the adjective. No suggested actions.

DPC [Phrase misselection] error and error of [Coherence]

Error 33. '*Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, *it is vein for the student to have so many good skills.*' Five participants considered this [Phrase misselection] error an error of [Coherence]. It is understandable that this was categorised as incoherent, however, most participants did manage to identify that it was deviant of 'it is in vain'. Some seem to understand intended meaning well and others do not. It could also have been classified as a [Spelling] error. See discussion under DCPM. No suggested actions.

DPVE [Phrase misselection] error and error of [Verbosity]

Error 18. '*So, what score can a student obtain is determined by *how to organise words to indicate points.*' Two participants considered this [Phrase misselection] error as an error of [Verbosity], which is understandable: a better reformulation would have been '<...*determined by how to structure sentences*'. This appears to be problematic: different analysts will reformulate errors differently, and from that, categorise the errors in different ways. The solution is to reiterate that the error should be categorised based on what was written on the page, and not the reformulation.

Coherence

DCU [Coherence] error and error of [Underspecification].

Error 34. '*Speaking to others and get more information from others is a correct way to *study academy and improve one's ability.*'

Error 36. '*I think, the most important is not only reading skill. *Other skills are not. either .*'

Eight participants considered these [Coherence] errors to be errors of [Underspecification]. This is completely understandable as adding more and different words would indeed enable the student to better express herself. Some participants understood the sentence, others did not. One cannot cater for different

participants' willingness or ability to understand or make them follow instructions to check the whole essay to see errors in context to assist with understanding.

DCS [Coherence] error and error of [Suffix omission]

Error 34. '*Speaking to others and get more information from others is a correct way to *study academy and improve one's ability.*' Three considered this [Coherence] error to be an error of [Suffix omission]. This is also completely understandable, as writing 'academically' instead of 'academy' would have improved accuracy, but again, there is little to be done if some participants understand the sentence and others do not, especially after they had been asked to refer to the essay to see the error in context, which might have helped with understanding.

DCPM [Coherence] error and error of [Phrase (misc) disorder]

Error 35. '*I think, *the most important is not only reading skill. Other skills are not. either .*' Three participants considered this [Coherence] error to be an error of [Phrase (miscellaneous) disorder]. Certainly, reordering the words in this phrase would make it clearer. Again, some were able to guess at an intended meaning and others were not.

Reducing the word count in the guidance and making instructions clearer may help to reduce many of the errors above. However, difficulties are envisaged when expecting all analysts to comprehend and interpret the intended meanings of the errors in student writing.

Noun overinclusion

DNN [Noun overinclusion] error and [Noun misselection] error

Error 6. '*Reading *skill, undoubtely, is a very crucial factor to master.*' It is unclear why two participants saw this as a [Noun misselection] error: no action. On reflection, there is little wrong with this 'error'. It was included originally as I felt that omitting the word 'skill' would have been better. It seems that the inclusion of these 'less serious' errors reduced overall agreement scores and increased DCIs.

DNB [Noun overinclusion] error and error of [Two words should be one (blend)]
Error 6. '*Reading *skill, undoubtely, is a very crucial factor to master.*' Two participants saw this [Noun overinclusion] error as a [Two words should be one] blend error. Despite James' (1998) sensible suggestion for adding [Blend] to the surface error taxonomy to ensure there are specific categories for all types of errors, there do not appear to have been many [Blend] errors. In fact, the [Blend] category in this research so far has only been used by participants who categorise errors in a way that is different from the mode. Removing the [Blend] column would address four of the errors above and reduce the number of categories in the framework and word count in the guidance. Suggested Action: remove [Blend] from the horizontal axis.

DNM [Noun misselection] error and error of [Single word (miscellaneous) misselection]

Error 8. '*Reading skill, undoubtely, is a very crucial *factor to master.*' Four participants considered this [Noun misselection] error an error of [Single word (miscellaneous) misselection]. Removing the [Single word (miscellaneous)] category would address 19 DCIs. It was a flaw in the design to have included this in NewLEAF2. Other word class categories should account for all single word errors.

DNU [Noun overinclusion] error and [Uncategorisable] error

Error 6. '*Reading *skill, undoubtely, is a very crucial factor to master.*' Two participants saw this [Noun overinclusion] error as an [Uncategorisable] error. Perhaps they did not see the inclusion of 'skill' it as an error and felt they had no choice but to allocate it to [Uncategorisable] (see discussion under DNN).

Spelling

DSSO [Spelling misselection] error and error of [Spelling overinclusion].

Error 13. '*different kinds of books *essaies, files and so on.*' Three participants considered this [Spelling misselection] error an error of [Spelling overinclusion]. This is understandable as the error seems to be both misselection (of plural affix) and overinclusion of more letters than are necessary, but the more specific and

helpful error explanation is the former. The guidance states that 'overinclusion' meant that the issue is that there is an extra letter or word/words that does not belong, which is not very clear. Suggested Action: clarify this in the guidance and framework.

DSSM [Spelling misselection] error and error of [Suffix misselection]

Error 13. '*different kinds of books *essaies, files and so on.*' Three participants considered this [Spelling misselection] error an error of [Suffix misselection]. The guidance states that plural prefixes are to be excluded from the analysis. Perhaps this level of detail was too much for some. Suggested action: simplify the guidance and framework and reduce overall wordcount.

DSS [Spelling] error and error of [Suffix omission]

Error 2. '*To get a perfect academic *performan is a basic quality that every student wants.*' Three participants considered this [Spelling] error an error of [Suffix omission], which is odd as the root here is perform, not '*performan*'. This points to DCIs that could be caused by analysts' limited linguistic metalanguage. There were six DCIs involving [Spelling] and [Suffix] categories. Suggested Action: make [Affix] errors sub-sections of [Spelling] and ask analysts to allocate spelling errors within affixes to the [Affix] categories. This would add systematicity.

DSSB [Spelling misselection] error and [Spelling blend] error

Error 25. '*content of what prof*fesser says clearly.*' Two participants saw this [Spelling misselection] error as a [Spelling blend] error because there are two spelling errors: double f and -er ending (instead of -or). See discussion and suggested action under DNB.

Verb misselection

DVP [Verb misselection] error and [Preposition omission] error.

Error 32. '*Whatever a student hear and write and how fast a student read, if a student can not *talk own views with others, it is vein for the student to have so many good skills.*' Two participants saw this [Verb misselection] error as a [Preposition omission] error. It is not clear how inserting a preposition before or after the erroneous '*talk*' would have provided a suitable fix: no suggested action.

DVS [Verb misselection] error and error of [Single word (miscellaneous) misselection]

Error 24. '*Before you use the new knowledge, you need *hear the content of what professor says clearly.*' Two participants considered this [Verb misselection] (should be 'understand') error an error of [Single word (miscellaneous) misselection]. Suggested Action: removing the [Single word (miscellaneous) misselection] category will stop DCIs with [Verb misselection].

Two words should be one

DTU [Two words should be one] error and [Uncategorisable]

Error 31. '*Whatever a student hear and write and how fast a student read, if a student *can not talk own views with others, it is vein for the student to have so many good skills.*' Two participants saw this [Two words should be one] error as [Uncategorisable]. This seems to be a case of participants not considering this an incorrect form, or not being aware of the [Two words should be one] category. Suggested Action: reduce the detail in the guidance and framework. Simplifying the guidance and framework would place less of a burden on analysts' short-term memory and may enable them to see or remember appropriate categories (they cannot see the wood for the trees).

Other researchers have found categorisation problematic when conducting LEA. Doolan and Miller (2012), for instance, found that it was difficult to differentiate between idiomatic expressions, collocations and prepositional phrase errors and therefore grouped these into a single [Prepositional phrase] category, as it is 'often the token signalling this language problem' (Doolan and Miller, 2012:12). They also stated that they had to exclude certain categories from their analysis as they could not be reliably allocated to. These included [Oddly constructed phrases], [Idiomatic expressions], [Fractured syntax], and [Coherence breaks]. They added that 'if defined reliably, [they] might uncover further valuable insights into student error patterns that were not measured in the current study.' Doolan and Miller added that not all of the errors identified in their study were categorised, however, the nine categories that were used accounted for approximately 75–85% of total errors, which points again to the imperfect, yet still useful nature of EA.

It was often felt that if the participants had been concentrating a little more or had understood or not been confused by the amount of details in the guidance and framework, they may have categorised in line with the mode a little more. Indeed, there were several instances where I could not understand why some participants had categorised in the way that they had. These may appear to be instances where the participant was not concentrating nor following instructions. However, the participant should not be blamed for these inconsistencies. Rather, simplifying and clarifying the guidance and framework would make it easier to use.

The above suggested actions have been included to amend the guidance and framework (NewLEAF3), which can be found in Appendix 7.16.

7.4 Conclusion

This section summarises the issues encountered with, and offers further discussion on, identification and categorisation. It also discusses perceived limitations with the current study.

7.4.1 Identification

It is likely that identification of LEs is so problematic that regardless of measures taken to bring thinking into line, inter-rater agreement will continue to vary considerably. Zimmerman (1987) discusses issues of error identification and shades of acceptability. James (1998) also discusses awkward expression: language that is not so much wrong, but perhaps inappropriate, or not what a highly-skilled language user might use and recommends inclusion of these instances in EA. However, opinions on what is considered an error or simply awkward expression are subjective and will therefore inevitably vary between analysts. Differing views on strictness will also reduce inter-rater agreement. Perhaps the low agreement score for error identification could also be explained by the added difficulty of being able to satisfactorily differentiate between lexical and grammatical errors. Therefore, perhaps it is to be expected that there is some variation between analysts: it cannot be completely removed, but it could be reduced. It would be very difficult, contentious and time consuming to implement sufficient advice into the guidance that would allow analysts to reliably identify

LEs. Further, it has been shown that this study not only tested the framework and guidance, but also participants' abilities to follow instructions, concentrate, notice errors, identify them as such and use the tools. The design of the framework and guidance and the analysts' abilities to use them can be seen as two separate entities. This said, this low, overall, inter-rater agreement score does not promote the use of LEA. However, it can be said that many individual errors did attract high agreement scores within certain sub-groups of participants and that LEA should not be dismissed on these grounds as a useful analytical tool for individual teachers wishing to systematise LE feedback to individual learners or groups.

It seems that despite the provision of a definition in the instructions, agreement of identification and categorisation of LE is still problematic for the majority of raters in the current study. Whilst the finding that raters demonstrated much difference in their error identification was a disappointing one which did not support the call for LEA to be more widely adopted, it is nonetheless an interesting and potentially important finding: this great variability in LE identification needs to be addressed, if only for the better standardisation of high-stakes testing. IELTS and a number of other established proficiency tests, such as the Cambridge suite of exams, have a lexical grading criterion. If marking is to be accurately normed, then better guidance is required to help raters agree on what constitutes an LE, not to mention guidance on what constitutes a more serious LE that will affect a rater's grade more severely than another.

One could further speculate as to the reasons behind the low identification agreement scores across both groups. I suspect that the nature of error identification might be highly subjective and depends on a variety of factors, as discussed in Section 7.3.2.1, making it inevitable that there would be differences among and between groups. There does not appear to be a quick fix solution to modifying the guidance to address this issue. In order for raters to identify errors consistently, they would have to have a shared understanding of which items are 'correct' and which are 'incorrect'. Unfortunately, this simple difference is not reflected in language. Instead, we seem to have 'correct, acceptable, slightly acceptable and erroneous' language, and different raters with different standards and different views and tolerance of errors will continue to use their own

judgement, based on their own backgrounds and understanding of rules when identifying LEs in different modes and genres. The problem is possibly greater for English because its status as a world language implies greater variation in usage. This great subjectivity is proved by the fact that each participant identified at least one LE that the others did not.

On the matter of whether to count repeated error tokens in an LEA, Hemchua and Schmitt's 2006 study was an investigation into error types; perhaps they felt that counting repeated tokens would exaggerate the frequency of errors, and identifying true identical errors (i.e. cases where the erroneous and likely target forms had to be identical), could prove to be problematic. They therefore ignored repeated tokens. On reflection, it seems preferable to me to include all repeated lexical error tokens, because omission of duplications would hide how widespread the issue was. Also, it was felt to be more logical to include all errors, if they were made. This would also circumvent the earlier discussed issue of differentiating between types and tokens and also between error and mistake: inclusion of all would simplify matters.

These details point to the rather subjective, unstable, chaotic and arbitrary nature of LE identification. It could be that the learning curve required for LE identification for inter-rater reliability is too long and too steep to make LEA practical.

7.4.2 Categorisation

Interestingly, the mode of the categorisations, found by comparing the categorisations of all the experienced, qualified English language teachers who had English as a first language, matched my categorisations in all of the 37 cases except three. This argues that there is promise of good agreement for categorisation when the framework is used by this group. Several changes were made to the framework following this study: the main ones were the removal of the [Single word (misc) misselection] category, the [Blend] column and the reorganisation of some categories as subcategories of others. It is hoped that this will guide future analysts to allocate to the more specific subcategory, which will increase agreement scores. One of the main areas that involved DCIs was that of

[Formality]: some participants, who could not identify formality errors as such, allocated them to the [Uncategorisable] category. As stated earlier, it would be tempting to remove the [Uncategorisable] category in an effort to make analysts think again and hopefully increase agreement. However, retaining this category allows for a 'bin' where errors that do not fit into the existing categories can be placed for reconsideration of the design of the framework. Perhaps it is a category that could have been excluded for the trialling, but retained in future versions. This could be seen as a limitation of the last two studies, and this and other limitations of the current study are discussed in the next section.

There are now 13 fewer categories in the framework, down from 56 to 43. Further, the total word count in the guidance and framework has been reduced from 1,560 in NewLEAF2 to 1,213 in NewLEAF3. It is hoped that by reducing choice and clarifying the guidance and framework by reducing and clarifying information, greater agreement can be achieved.

Despite the fact that there was 80% agreement in Part One and Group A's agreement score was an encouraging 64%, categorisation still does not appear to be systematic enough for reliable findings in studies conducted by multiple analysts. However, this does not detract from its potential use by single researchers or teachers who can analyse consistently in line with their own system. It is hoped that further amendments to NewLEAF3 will result in improved agreement between analysts in future.

7.4.3 Limitations of Study 5

The participants' ability to identify LEs was not established before asking them to complete Task 1. It could be the case that some participants could differentiate between lexical and grammatical errors better than others following training. It would have been useful to only use such participants. However, in the previous study, only six participants were used. These were used again in the current study for variable control purposes, such as ability to follow instructions. Results using these participants only would not have been significant. Further, whilst some errors are clearly wrong, interpretations of lexical correctness appears to be subjective, so there would not have been a correct set of correct lexical usages to compare their identifications against.

Perhaps it is unsurprising that the six participants used in both the current study and previous Study Six had a higher agreement score in terms of categorisation than the rest of Group A and Group B. Perhaps this is due to the fact that this would have been their second attempt at LEA, which points to the possibility of greater agreement coming with familiarity: a sentiment stated by participants in the previous study. It could also have been due to the fact that the six participants were colleagues of mine, were also in the same room as me, and may have therefore tried their best to follow instructions, whereas others may not have been able to perform as well.

There is an issue with extent of error identification. When calculating agreement scores for error identification, I counted an underlined error with a single word as the same error as an error that was identified as a phrase that included that word. Had these errors been counted separately, the identification agreement scores would have been even lower. Therefore, more guidance on extent of error needs to be included in the guidance.

Unfortunately, data collection took place during the height of the Coronavirus pandemic, a time when language teachers were busy moving their instruction online. Contacts were asked to forward the call for participants to their contacts. This may have been done, but unfortunately, the intended snowball sampling did not occur. Perhaps these factors explain the low response rate.

Only one essay was used in the current and previous studies. However, there are reasons for this: it was produced by someone whose first language is very distant from English (Chinese, so it is more likely that this learner would produce more erroneous language than a learner whose L1 was closer to English), it includes a wide range of LEs, but more importantly, the framework had to work well on one essay before it could be tested on essays from a whole cohort and essays from students of different L1 groups. Perhaps future studies could focus on investigation into the type and frequencies of LEs from a whole cohort of students or a group of students from one L1 group so that the results could be compared against LEs from another language group.

This study has sought to produce a better version of the guidance and framework. It is acknowledged that NewLEAF3 has not yet been tested for inter-rater agreement. Unfortunately, the scope of this study does not allow for further testing and refinement, and the cycle of testing and refinement must be stopped at some stage. NewLEAF3 is as good as it could possibly be in light of some of the prevailing issues with EA, namely, error identification, separation and categorisation. Given these problems, it is highly unlikely that a perfect system that will allow for 100% agreement in these areas can be found. That said, this final, current study uncovered only 112 (out of a possible 666, 17%) instances of dual categorisation issues that could be addressed without reducing the total number of categories and therefore the potential usefulness of the framework. A system of categorisation that offers a better agreement score and very few possibilities for dual categorisation could be devised, but this will involve the sacrifice of further categories, such as [Formality], [Underspecification] and [Verbosity], leaving only [Phrase] error in Section B, but this would be of limited use to teachers and students. A useful balance between ease of use and detail of results, perhaps, has been found.

Despite the seemingly serious set of limitations laid out above, inter-rater agreement of identification and categorisation is only relevant when teachers or researchers would be using the system to compare results of different groups or individuals, or the same groups or individuals when different analysts are used. Intra-rater agreement is a separate issue. The framework is probably still perfectly adequate for most purposes, since most users will be working alone with one language group for their own individual purposes – i.e., working with their class of students, as long as he or she can be internally systematic. However, this has not yet been tested and is a possible direction for further research.

In Study 6, NewLEAF3 is applied to the same 20 scripts, written by Greek learners, that were used in Study 1. It is important to use the same data so that comparisons can be made between Hemchua and Schmitt's (2006) framework and the latest version. This will enable conclusions to be drawn in terms of ease of use and depth of analysis and the results themselves will also be of interest.

Chapter 8 Study 6 Implementation of NewLEAF3

8.1 Introduction

NewLEAF3 was designed following an investigation into the ease of use and depth of analysis of existing LEA frameworks (Study 2). It was further refined in Studies 3 and 4 through qualitative investigation into ease of use and how it may be integrated into the work of practising teachers, and also through quantitative investigations into agreement of error identification and categorisation between participant analysts. These investigations and subsequent refinements produced NewLEAF2 and 3. This study seeks to test the latest refinement, NewLEAF3, using the same set of essay data that were used in Study 1 to establish what types of LE and their frequency would be found when using the new version. Direct comparison of the results with Study 1 would be problematic, as the two frameworks employed different taxonomies of error and guidance in terms of counting repeated tokens, but some comparison is possible. Furthermore, it is important to establish how easy it is to use NewLEAF3 and how satisfactory a spread of results it produces in comparison to a previous framework and guidance (Hemchua and Schmitt 2006). Therefore, the research questions are as follows.

8.2 Research questions

- 1) What lexical error types and frequencies are found when NewLEAF3 is used on a set of student compositions?
- 2) How does the depth of analysis and spread of results compare to those produced by Hemchua and Schmitt's (2006) framework in Study 1?
- 3) How easy to use are the guidance and framework in comparison with Hemchua and Schmitt's (2006)?

8.3 Methodology

LEA was performed using NewLEAF3 on the same 20 essays produced by Greek learners that were used in Study 1 (see Study 1 for a description of the participants, ethical consideration, how the data were collected and methodology). It was important to use the same essay data, as this would enable better comparison between the two frameworks and guidance in terms of depth of analysis and ease of use. The only difference in methodology was that repeated

tokens were included in the error count and confidence scores were attached to each categorisation decision in Study 6, as they were in Studies 3 and 4, so that my overall sense of categorisation surety could be established. The analysis was also timed to broadly establish whether LEA was time-consuming or not.

8.4 Results and discussion

This section is organised by research question.

Research question 1 - What lexical error types and frequencies are found when NewLEAF3 is used on a set of student compositions?

As can be seen from Table 45 below, there were 299 LEs found in the 20 scripts

Error type	Omission	Over-inclusion	Mis-selection	Misorder	Totals	% of all errors
A1 Suffix	5	2	12	N/A	19	6
A2 Prefix	2	0	4	N/A	6	2
A3 Spelling (misc)	5	0	0	1	6	2
A4 Two words should be one	4				4	1
A5 One word should be two	0				0	0
B1 Coherence	17				17	6
B2a Conjunction	4	7	2	0	13	4
B2B Noun	1	22	17	N/A	40	13
B2C Adjective	0	1	5	0	6	2
B2D Adverb	N/A	1	5	N/A	6	2
B2E Preposition	2	4	35	N/A	41	14
B2F Pronoun	2	0	5	0	7	2
B2G Verb	5	3	30	0	38	13
B3 Phrase (misc)	N/A	N/A	74	3	77	26
B3A Misordering in fixed phrases	0				0	0
B3B Verbosity	4				4	1
B3C Underspecificatio n	8				8	3
B3D Formality	7				7	2
B4 Uncategorized	0				0	0
Totals					299	100

Table 41 Results of LEA on Greek Data using NewLEAF3

Across 16 error types. 37 (12%) were formal errors and 264 (88%) were semantic. The most common of the latter were [Phrase] errors: 77 in total, accounting for approximately a quarter of all errors. It is perhaps unsurprising that this was the largest category, as it acts as a catch-all category for multiple errors in a phrase that included non-standard lexical choices where the error spanned across several words and consequently could contain a variety of word class types. Errors in [Prepositions] were also very common (14% of the total), which is unsurprising, given the comparative lack of meaning that dependent prepositions carry: in my experience, this is the most common type of error made by more advanced learners. The next most common error category was errors with [Verbs] (13%) and [Nouns] (13%). There were 17 (6%) errors of [Coherence], which I consider the most serious error type as they cause breakdown in communication. Next, there are errors with [Suffixes] (6%), the most common formal error type. There were very few [Spelling] errors. Three categories were not represented at all: [A5 One word should be two], [B3A Misordering in fixed phrases] and [B4 Uncategorisable].

Research question 2) How does the depth of analysis or spread of results compare to those produced by Hemchua and Schmitt’s (2006) framework in Study 1?

It is interesting to compare the overall results below in Table 46 with Study 1, as the same essay data were used. However, as different categories were used in the two frameworks, comparison of specific error types is, of course, problematic.

Study	Total word count (20 essays)	Mean word count	Stan Dev	Min	Max	No LE Tokens	Av No LE Tokens per paper	Error Types
1	5,912	295.6	47.16	178	407	284	14.20	16
6						299	14.95	16

Table 426 Comparison of lexical errors found using Hemchua and Schmitt’s (2006) framework and NewLEAF 3 when using the Greek data

Error count

As shown in Table 47 below, 299 LEs were identified with the new framework, whilst 284 were identified in Study 1, where Hemchua and Schmitt’s (2006) framework was used. In Study 1, there was an average of 14.2 errors per paper. In

Study 6, there was an average of 14.95 errors per paper. The similarity in figures, a negligible difference of only 15 LEs identified in total, points to good intra-rater consistency in terms of error identification. This may be a consequence of using the same data and similarity between Hemchua and Schmitt's (2006) methods of identification and those used in NewLEAF3. There were the same number of error types in both studies, despite the fact that the more recent study did not attempt to include causality categories.

Types of errors made- formal and semantic errors

Study	Formal Errors		Semantic Errors	
	No of Errors	% of total errors	No of Errors	% of total errors
1	82	28.87	202	71.13
6	35	11.71	264	88.29

Table 437 Frequency of formal and semantic errors in Study 1 and Study 6

As shown in Table 47 above, there is some similarity in the distribution of the type of error in terms of formal vs semantic between the two studies. In both, the majority are semantic. However, the greater percentage of formal errors in Study 1 is accounted for by the fact that there are extra categories in the formal section in the Hemchua and Schmitt framework, including several for [L1 interference errors]. For logistical and accuracy reasons, causality was, as discussed in Section 2.4.5, omitted from the new framework.

Problems with formal errors

As shown in Table 48 below, the most frequent formal error type in Study 1 was [A2.3 Calque (translation)], (12% of total errors), followed by [A1.1 Suffix Type] (9.2%). In Study 6, they were [A1 Suffix] (6%), followed by [A3 Spelling (2%)]. This difference is due to the fact that NewLEAF3 does not attempt to categorise by causality.

Problems with semantic errors

In Study 1, the most common semantic error types were [B2.1 Semantic word selection] (31% of all errors), then [B2.4 Preposition partners] (18%), followed by [B1.4 Near synonyms] (10%). In Study 6, they were [B3 Phrase error] (26%), [B2E Preposition] (14%) and [B2G Verb] (13%) and [B2B Noun] (13%). Similar figures

can be seen for errors with prepositions, which of course is due to the fact that the same essays are used: both frameworks have a [Preposition] category and classification of errors with prepositions is relatively unproblematic. The most common LE across the two studies is misselection of individual words.

Overall

The addition of the surface taxonomy, which was placed in the horizontal axis, provides for more depth of analysis as there are now 43 different specific types of error categories (including surface taxonomy sub-categories), as opposed to 23 in Hemchua and Schmitt's (2006) framework. In terms of actual results, Table 46 below shows the rankings of most common individual error types found using the two different frameworks. As can be seen from the results, NewLEAF3 offers a slightly better spread of results, even without error type categories for L1 interference, as used in the Hemchua and Schmitt (2006) framework. There is less grouping in NewLEAF3 in the most common eight categories than there is in Hemchua and Schmitt's framework, and the errors are spread over more categories in total in NewLEAF3. In Study 5, errors were found in 16 of the 19 available categories (84%). This provides a more satisfactory spectrum of error types, as it exposes a greater range of error types for the teacher and learner to understand. This alludes to the fact that Hemchua and Schmitt's (2006) framework did not account for or fully explain as many error types. However, as stated above, approximately a quarter of all errors were categorised as [Phrase] errors. This could be seen as over-grouping, to some extent, but given the difficulties associated in subcategorising this broad category into mutually exclusive subcategories, this figure can be seen as satisfactory.

With NewLEAF3, the top four categories accounted for 66%. After this, categories report single figures and no instances for three categories. In Study 1, errors were found in only 15 of the 23 categories (65%): one fewer than were found in Study 6 (84%). Over-grouping of errors was more of an issue in Study 1, as the four most common error types accounted for 73% in Study 1, 7% more than in Study 6.

Where there are the same categories in both frameworks, for example [Suffix] and [Prepositions], minor differences between reported numbers could be explained by dual categorisation issues in Study 1 where some errors were categorised under L1 interference categories.

Study 1				Study 6			
Ranking	Error Type	No Errors	% of errors	Ranking	Error Type	No Errors	% of errors
1	B2.1 Semantic word selection	89	31	1	B3 Phrase Error	77	26
2	B2.4 Preposition partners	51	18	2	B2E Preposition	41	14
3	A2.3 Calque (translation)	34	12	3	B2G Verb	38	13
4	B1.4 Near synonyms	29	10	4	B2B Noun	40	13
5	A1.1 Suffix type	22	8	5	B1 Coherence	17	6
6	B4.2 Underspecification	18	6		A1 Suffix	19	6
7	B4.1 Verbosity	13	5	7	B2A Conjunction	13	4
8	A3.1 Omission	12	4	8	B3C Underspecification	8	3
9	A3.4 Misordering	6	2	9	B3D Formality	7	2
10	A1.2 Prefix type	3	1		B2F Pronoun	7	2
11	B1.1 General term for specific one	2	1	11	A3 Spelling	6	2
12	A1.4 Consonant-based type	1	1		B2C Adjective	6	2
	A2.1 Borrowing (L1 words)	1	1	B2D Adverb	6	2	
	A3.2 Overinclusion	1	1	14	A2 Prefix	6	2
	A3.3 Misselection	1	1	15	A4 Two words should be one	4	1
16	A1.3 Vowel-based type	0	0		B3B Verbosity	4	1

	A1.5 False friends	0	0	17	A5 One word should be two	0	0
	A3.5 Blending	0	0		B3A Misordering in Fixed Phrases	0	0
	B1.2 Overly specific term	0	0		B4 Uncategorisable	0	0
	B1.3 Inappropriate co-hyponym	0	0				
	B2.2 Statistically weighted preferences	0	0				
	B2.3 Arbitrary combinations	0	0				
	B3 Connotative meaning	0	0				
Totals		284	100			296	100

Table 44 Types and frequencies of the LE s made in Study 1 and Study 6

As can be seen from the previous paragraphs, NewLEAF3 reports a slightly wider spread of results and also shows less bunching across the most common categories than Hemchua and Schmitt's (2006) framework, arguing that NewLEAF3 provides a better range of errors than Hemchua and Schmitt's and the former reveals some errors that the latter does not. Less use has also been made of the [Uncategorisable] category in NewLEAF3, which shows that NewLEAF3 finds classification for more errors.

Research question 3 How easy to use are the guidance and framework in comparison with Hemchua and Schmitt's (2006) when they were used in Study 1?

Overall, I felt that the NewLEAF3 guidance and framework were easy to implement consistently, as I believed that previous issues in error identification and, to a larger extent, in categorisation, had been overcome.

Error identification and count

In Study 6, as compared to Study 1, fewer issues were found in the identification of the LEs. As stated earlier, this seems to be an issue only when dealing with agreement between larger numbers of raters who had perhaps not familiarised themselves with the guidance to the same extent that the researcher had, and had to give consideration as to whether an awkward expression should be included. It is believed that this may continue to be an issue with inter-rater agreement, but should not be an issue with intra-rater agreement after some initial familiarisation, as I, and some participants in Study 2, feel that one soon builds up an internal system for applying one's own beliefs about language correctness with the framework. Furthermore, counting repeated tokens speeded the process up as one did not have to check back to see if an error had already been counted.

Grammatical Vs lexical error

As discussed in Study 1, Hemchua and Schmitt's (2006) rules for determining which errors are considered grammatical, and should therefore be excluded from the analysis, lacked clarity and scope: Study 1 showed that some of the rules were difficult to implement. The improved guidance used in Study 6 produced no instances of indecision over whether an error should be considered lexical or grammatical, which makes NewLEAF3 an improvement.

Error categorisation In terms of surety of the categorisation of identified errors when using NewLEAF3, I recorded no confidence scores of 1 (0%), 15 (5%) scores of 2 and 284 scores of 3 (94%). The average confidence rating of 2.9 per error is an improvement on Studies 3 and 4, but the improvement is perhaps unsurprising, given my familiarity with my own framework. Unfortunately, confidence scores were not employed in Study 1, but there were several instances of dual-categorisation issues in Study 1, as reported in Section 3.5.3. Removal of causality categories and placing the surface taxonomy in the horizontal axis greatly reduced the possibility for dual categorisation. Particularly pleasing in Study 6, was the complete lack of errors placed in [B4 Uncategorisable], which demonstrates that at least I could systematically allocate all identified errors to defined categories.

In Study 6, each essay took on average 7.5 minutes to analyse. With an average word count of 296 per essay, this amount of time was felt to be acceptable.

8.5 Conclusions

Summary of answers to research questions

- 1) **What LE types and frequencies are found when NewLEAF3 is used on a set of student compositions?** Results showed that of the 299 errors found in the 20 essays, the vast majority were semantic: 35 (12%) were formal errors and 264 (88%) were semantic. The six most common individual error types were semantic: [B3 Phrase errors] (26%), [B2E Preposition] (14%), [B2G Verb] (13%), [B2B Noun] (13%), [B1 Coherence] (6%) and [A1 Suffix] (6%). Of all error types, [Misselection] was the most common sub-category. Three categories, [A5 One word should be two], [B3A Misordering in fixed phrases] and [B4 Uncategorisable], saw no errors, suggesting that they are lower frequency errors, and that errors are not evenly divided across the spectrum.

- 2) **How does the depth of analysis or spread of results compare to those produced by Hemchua and Schmitt's (2006) framework in Study 1?** These results can be seen as satisfactory in terms of depth of analysis because in comparison with Study 1, a slightly wider range of errors were reported in Study 6. They were also spread more evenly across the taxonomy than was seen when using Hemchua and Schmitt's (2006) framework. NewLEAF3's results saw less over-grouping of errors across both formal and semantic error categories and can therefore be considered to be more accurate and efficient in LEA as previous versions saw a mixture of error types in fewer categories whereas NewLEAF3 separates these more so that they can be better analysed. Furthermore, the five most common error types accounted for 79% in Study 1, whereas in Study 6, the figure fell to 72%. Interestingly, where comparison was possible, the results from Study 6 were quite similar in places to those found in Study 1, despite the fact that there were some different error type categories in the two frameworks.

3) How easy to use are the guidance and framework in comparison with Hemchua and Schmitt's (2006) when they were used in Study 1?

NewLEAF3 was relatively straightforward to use as there were no issues in error identification, nor decisions over whether an error could be viewed as lexical or grammatical in nature. However, issues in the latter were found in Study 1. When using NewLEAF3, errors could be allocated to categories confidently and no errors were found to be uncategorisable, which suggests that the framework is easier to use than the one published in 2006.

Implications

These results have implications for the teaching of lexis. The greater number of errors made were in the [Phrase] and [Preposition] categories (40% of error types), which provides strong indirect support in favour of a lexical and collocational approach to language teaching, with more focus on developing student knowledge of suffixation and word families. These implications reflect those found by Hemchua and Schmitt (2006).

Limitations of study

The analysis in this study was conducted by myself. Therefore, there could have been some overconfidence in allocation of confidence scores due to my own high level of expertise and familiarity with this framework. However, it was in my interest to produce a workable framework, so the analysis was conducted with integrity. That said, different results may have been produced by a different analyst who was less familiar with the framework.

NewLEAF3 was used on the same essay data (20 scripts from Greek learners) that were used in Study 1. Whilst this enabled comparison of results with Hemchua and Schmitt (2006), it did not enable identification of possible new error types. It did enable conclusions to be drawn in terms of ease of use and depth of analysis. The results themselves will also be of interest.

It would have been useful to have tested NewLEAF3 with a wider range of analysts with the same essay data. This would have enabled investigation into inter-rater identification and categorisation agreement.

Although it is slightly beyond the scope of this dissertation, the earlier claim that intra-rater reliability would improve with increased usage of the framework, as suggested by some participants in Study 6, was partially investigated. Five years after performing LEA on the Chinese student's essay in Appendix 6.6, I used NewLEAF to repeat the analysis on the same essay. I found 37 errors in each analysis. However, six were different errors from the original study. There were also differences in the extent of error identification. Overall, this represents an 84% similarity between the analyses. When the original 37 errors were categorised, 34 were categorised in the same way, giving a 92% similarity figure for categorisation. It could be argued that these promising results are unsurprising given that I have worked closely with this data over the last few years, but there is no reason to suppose that familiarity with the tool could not be achieved by other users.

Overall, the results above are positive and demonstrate that the framework and guidance are satisfactory when used by a researcher wholly familiar with them. Conclusions, further limitations and future directions for LEA are outlined in the following chapter.

Chapter 9 Conclusion

9.1 Summary and conclusions

This chapter offers a summary of the studies and their findings. It also outlines the conclusions that have been drawn from the work and the limitations of the study as a whole and of the latest version of the framework and guidance. Finally, suggestions for further steps and research in this area are suggested.

9.1.1 Literature review

Chapter 2 examined what has been published in a range of subtopics related to EA and LEA. Section 1 of the literature review put forward the argument that despite the increasing popularity of fluency-based approaches to language teaching, there is still a need for error correction, and particularly for LE correction. This is largely driven by the requirement for lexical accuracy in high-stakes tests such as IELTS. Section 2 established ongoing limitations of computerised EA, which is still unable to accurately identify and categorise errors and requires continuation of manual tagging. Section 3 provided some historical contextualisation by describing the rise and fall of EA through the 1960s and 1970s. EA's issues were procedural in nature and there were questions regarding its suitability as it became associated with the now largely discredited teaching methodologies of grammar translation and audiolingualism. It was suggested that some of the procedural issues could be overcome, but this required further investigation. Overall, EA as a tool for analysis of learner writing should not be abandoned. Section 4 took a closer look at the stages of EA (collection, identification, description, explanation and evaluation) and examined their problems. The two main stages of identification and categorisation are the most important for successful EA, and are also the most problematic. A rebuttal of some of the limitations was put forward and the argument presented that for LEA to be of use, the original EA procedures required some modifications, and that it remained an important tool for investigating learners' interlanguage development. Section 5 focused on LEA and argued for the centrality of lexis in language teaching and learning. It discussed a relatively recent view of lexis and grammar, lexicogrammar, and concluded that the two are difficult to separate for analytical purposes. The section also reviewed previous attempts to conduct LEA and the

potential of various approaches and frameworks in terms of ease of use and resultant depth of analysis. The section concluded that there is no perfect taxonomy for categorisation of LEs. It seemed that there is a balance to be struck between ease of use and depth of analysis: too many categories could lead to dual-classification issues and too few leads to a restricted depth of analysis. Overall, it would be worthwhile to establish whether an improved framework for LEA could be designed that at least mitigates some of the issues described above, even if there is still some overlap in the identification and categorisation between the types of lexical errors that can be made, as this would provide more comprehensive understanding of type and frequency of errors made.

9.1.2 Study 1

This was a replication study using Hemchua and Schmitt's (2006) framework and guidance, which seemed to offer the best balance between ease of use and depth of analysis, to analyse the type and frequency of LEs in the compositions of 20 Greek learners. The results were remarkably similar to those found in Hemchua and Schmitt (2006), despite the fact that gender, age and proficiency level were not controlled, and a different L1 group was used. Approximately two-thirds of errors in both studies were semantic in nature and one-third were formal. Furthermore, there was similarity in the total number of errors made and it was also found that [Near Synonyms], [Prepositional partners] and [Calque (Translation)] categories of error types appeared in the top four rankings in each study. This supported Hemchua and Schmitt's (2006) suggestion that learners with other L1s would probably make similar errors. It also suggests that the 2006 framework is relatively fit for purpose, and that it can be used with different L1 groups. There were, however, some issues found during the analysis, namely, difficulties in deciding whether a clause contained an error, whether that error should be considered grammatical or lexical, and if the latter, to which category it should belong. Suggested improvements included clearer guidelines on what to exclude in terms of grammatical items, clearer guidelines on allocation of lexical errors to category when there is more than one possibility: separation of cause and type of error and further sub-categories that will allow for categorisation of all errors (e.g., coherence, cohesion, awkward expression, missing word, and more

precise allocation of sub-types of errors. (i.e., whether some error categories, e.g., a preposition partner error, is an omission, addition or substitution error)).

9.1.3 Study 2

Five of the Greek essays were used to analyse the advantages and disadvantages of six different previously published LEA frameworks, selected for their different approaches. Whilst some were found to be easier to use and provided more depth, others yielded fewer uncategorisable errors, but none were found to be perfectly fit for purpose. There was some variation in terms of the numbers of errors the different frameworks revealed, due to the different ways in which the original authors classified LEs. Despite these differences in the numbers of errors found, they were quite similar: Zimmerman's (1986) and James' (1998) frameworks yielded the most errors (116 each) and Dušková's (1969) the fewest (53). This is quite surprising, given that Hemchua and Schmitt's (2006) had many more categories. There were also great differences in how the errors were categorised: the ease of use and the depth of analysis that they provide and also in the number of uncategorisable errors that they produced: Dušková's (1969) yielded the most uncategorisable errors (57.3%) and Corder's (1973) the fewest, yet both had a similar number of categories, suggesting that it is not the number of categories that is important, but the type. James' (1998) framework yielded only 11 uncategorisable errors, yet had many more categories, suggesting that his framework offered the best balance in terms of ease of use and depth of analysis. It was found that a common and useful method of categorisation is formal and semantic errors. It was also felt that removing cause from the analysis reduces speculation and issues with dual categorisation and, following James (1998), having two axes for categorisation (type of error and the surface taxonomy of [Misselection], [Over-inclusion], [Omission], [Misorder] and [Blend] supplies a much more detailed picture than using only one of these dimensions. Categorisation by word class, a more mutually exclusive taxonomy, could result in fewer uncategorisable errors, but the value of categorising errors in this way may be of less value to the teacher and student than categorising them by LE type (e.g. [Verbosity], [Prepositional partner], [Completely wrong word (field)] and [Wrong word (sense relations)], etc). It was also felt that combining spelling errors into one category would simplify matters without reducing depth of analysis. Much was

learned from these analyses to produce a new framework with hopefully a satisfactory balance between ease of use and depth of analysis that would yield fewer uncategorisable errors.

9.1.4 Study 3

NewLEAF1 was described and trialled using five essays. It was found to be an improvement on previous attempts as it was made easier to use, provided more detailed guidance, did not speculate on causality at the categorisation stage, and employed two axes which would allow for further sub-categorisation in terms of [Omission], [Misorder], [Misselection] and [Overinclusion] as well as more specific error types. Further, it incorporated what were thought to be more mutually exclusive categories. Guidance for analysts was provided in identification, counting and categorisation of LEs. It was then tested using the same data from Greek learners that were used in Studies 1 and 2, for comparative purposes. The framework yielded a similar number of errors and no uncategorisable errors. There were also no instances of possible dual-categorisation. As expected, with a different set of categories, the types of errors found were different from previous analyses, which limits the possibility of comparison of findings.

9.1.5 Study 4

The aim here was to measure inter-rater agreement and gain insight into what six highly-experienced and highly-qualified EAP tutors thought about the framework and guidance. This was done by asking them to use it to perform LEA on the same single essay produced by a Chinese learner. In terms of similarity of results, there was 59.3% agreement in terms of error identification, and an inter-rater agreement categorisation average figure of 67%. The low agreement figure for categorisation was largely due to dual-categorisation issues which, in turn, were caused by too much similarity between categories, such as [Field Error] and [Completely wrong word]. The guidance was found to lack clarity in places, often due to the amount of information and the terminology. However, the participants felt quite confident about their categorisations. Participants stated that it was an improvement on the marking symbols system that they used for feedback on written work, it was more thorough, and they found it easy to use. They also stated that they saw value in

using it periodically on the whole of a cohort's work to establish type and frequency of LEs, which would inform remedial teaching. They offered useful feedback in terms of how the guidance could be improved (clearer use of grammatical metalanguage, for example). Data from the identification and categorisation tasks and the semi-structured interviews were used to produce NewLEAF2, which included clearer guidance and a new approach that separated the semantic part B of the framework more clearly into errors with single words and errors with phrases. The former incorporated a word class categorisation approach as this mutual exclusivity approach would lead to greater inter-rater agreement.

9.1.6 Study 5

This was also a test and refine exercise using the same data and procedures as the quantitative part of Study 3. NewLEAF2 was also tested for inter-rater agreement using a wider group of qualified teachers and students in linguistics and TESOL (38 in total). There was a mixture of participants who had English as their L1 or L2, with varied professional experience. Part A compared results from the same six participants in the previous study. Agreement scores for identification dropped to 51.36% overall, but rose to 80.11% for categorisation for this small group compared with results from Study 4, suggesting that categorisation with NewLEAF2 was easier to use when expert applied linguists with high-level language skills already had some familiarity with the process of LEA. However, the results from the group as a whole in Part B were less promising. Identification and categorisation agreement scores dropped to 18.69% and 55.32% respectively. These are discussed in more detail below.

9.1.6.1 Identification

When the overall results for all 38 participants were analysed, it was found that different participants had very different views in terms of error identification, as each participant identified at least one error that the others had not. It seems that there were highly subjective ideas about correctness in terms of [Formality], for example. It would be impossible to provide sufficient guidance about error identification to standardise this stage, as to cover all eventualities would be impractical. It is possible that identification of LEs is so problematic that regardless of measures taken to bring thinking into line, implying that inter-rater agreement

will continue to vary considerably, reflecting Zimmerman's (1987) point that issues of error identification are caused by shades of acceptability. James (1998) also discusses awkward expression: language that is not wrong but inappropriate, or not what a highly-skilled language user might use. In order for raters to identify errors consistently, they would have to have a shared understanding of which items are 'correct' and which are 'incorrect'. Unfortunately, this dichotomy does not exist in the English language. Instead, we have 'correct, acceptable and erroneous interpretations, and different raters with different standards and tolerance of errors will continue to use their own judgement, based on their own backgrounds and understanding of rules when identifying LEs in different modes and genres.

Perhaps the low agreement score for error identification could also be explained by the added difficulty of being able to satisfactorily differentiate between lexical and grammatical errors. Furthermore, it seems that despite the provision of a definition, the concept of LE is still quite vague to the majority of raters.

Whilst the finding that raters demonstrated much difference in their error identification was a disappointing one which did not support the call for LEA to be more widely adopted, it did unearth an interesting finding: this great variability in LE identification needs to be addressed, if simply for the sake of standardisation of high-stakes testing. IELTS and a number of other established proficiency tests, such as the Cambridge suite of exams, have a lexical grading criterion. If marking is to be accurately normed, then better guidance is required to help raters agree on what constitutes an LE and a more serious LE. This will affect a rater's grade more severely than another.

A more systematic approach to error identification might have been employed. For example, Doolan and Miller (2012) used prevalence and reliability of identification to identify errors. First, a random sample of eight texts were analysed by the researchers together for an initial set of error types for coding. 15 error types were found. The essays were re-read and recoded individually: the coding scheme was revised three times to produce nine categories. They assessed inter-rater reliability by first agreeing on the location of errors in 20% of the essays they examined and then coded them independently for error type. Inter-rater reliability was measured at 90%, which is a high percentage. However, this restricts the analysts to

searching for specific error types and is quite a small number of categories that would not give a wide range of errors or depth of analysis.

It seems a balance has to be found between sufficient guidance that will cover most issues in identification and too much information, which will prove difficult for participants who are new to the framework and guidance to work with.

9.1.6.2 Categorisation

Interestingly, the mode of the categorisations, found by comparing the categorisations of all the experienced, qualified English language teachers who had English as a first language, matched my own categorisations in all but three of the 37 cases. This augurs well for agreement about categorisation when the framework is used by this group. Although participants were quite confident about their categorisations, it was found that differences in agreement was caused mainly by similarity between [Formality] and [Phrase Misselection] and other categories, and also, it must be mentioned, by participants not following the instructions in the guidance and on the framework.

The potential for dual categorisation is reduced when the number of categories is reduced. The key is to have as many categories as possible without introducing the possibility for dual categorisation. Finding the right balance between ease of use (which includes reducing dual categorisation possibilities) and depth of analysis (having a broad spectrum of categories of error types, which facilitates a wider spread of results) has been an area of concern in the development of the new framework and guidance. Despite the fact that there was 80% categorisation agreement in Part One and Group A's agreement score was an encouraging 64%, NewLEAF2 still did not appear to be systematic enough for studies conducted by multiple analysts.

This study not only tested the framework and guidance, but also participants' abilities to follow instructions, concentrate, notice errors, identify them as such and use the tools. The design of the framework and guidance and the analysts' abilities to use them can be seen as two separate entities. That said, a low inter-rater agreement score does not support LEA. However, many individual errors did attract high agreement scores within certain sub-groups of participants, so LEA should not be dismissed as a useful tool for individual teachers wishing to systematise LE feedback to individuals or groups. However, it was clear that the

framework and guidance had to be reduced and simplified to increase inter-rater agreement, otherwise, the tool would not be suitable for comparison between studies.

9.1.7 Study 6

The final study sought to use NewLEAF3 on a set of essays produced by a single L1 cohort to ascertain the type and frequency of errors made by that L1 group. The original Greek data were used, as this would enable comparison in terms of ease of use and depth of analysis between Studies 1 and 6. It was found that the new framework offered a better spread of results, (more errors in a wider number of categories) and improved ease of use (fewer instances of dual categorisation).

In terms of results, of the 299 errors, 12% were formal errors and 88% were semantic. The five most common error types were all semantic: [Phrase] (26%), [Preposition] (14%), [Verb] (13%), [Noun] (13%), [Coherence] (6%). The joint sixth most common error type was formal, [Suffix] (6%). Of all surface error types, [Misselection] was the most common sub-category. Three categories of error ([One word should be two], [Misordering in fixed phrases] and [Uncategorisable]) were not found. In terms of depth of analysis, the results are satisfactory, as in comparison with Study 1, more types of errors were reported. They were also spread more evenly across the categories than in Study 1. NewLEAF3's results also saw less over-grouping across both formal and semantic error categories. Furthermore, the five most common error types accounted for 79% in Study 1, whereas in Study 6, the figure fell to 75%. Interestingly, where comparison was possible, the results from Study 6 were quite similar in places to those found in Study 1, despite some differences in the type of categories in the two frameworks. NewLEAF3 was easier to use than the 2006 framework as there were no issues in error identification, nor decisions over whether an error could be viewed as lexical or grammatical. Overall, errors could be allocated to categories more confidently and no errors were found to be uncategorisable.

9.2 Description and limitations of NewLEAF3

9.2.1 Description

The main improvements to NewLEAF3 (See Appendix 7.16) were that information relating to guidance in categorisation was included on the framework itself in terms of examples of errors so that rules could be more easily checked during the analysis, and the number of categories was reduced. The potential for dual categorisation was also reduced by introducing word class categorisation and fewer but clearer rules when potential dual-categorisation possibilities arose. Several other changes were made to the framework following the last study: the main ones were the removal of the [Single word (misc) misselection] category, the [Blend] column and the reorganisation of some categories as subcategories of others. It is hoped this will guide future analysts to allocate to the more specific subcategory, which will, again, hopefully, increase agreement scores. One of the main areas for dual categorisation involved [Formality]: some participants, who could not identify formality errors as such, allocated them to the [Uncategorisable] category. It would be tempting to remove the [Uncategorisable] category in an effort to make analysts think again and hopefully increase agreement: however, retaining this is useful when considering changes to the framework. Perhaps it is a category that could have been excluded for the trialling, but retained in future versions. There is no longer a [Collocation] category, which is a pity, given the current focus on multi-word unit teaching. However, retention would have led to too many dual categorisation issues. There is, however, a category for [Phrase] with several subcategories. NewLEAF3 has 13 fewer categories (down from 56 to 43). Further, the total word count in the guidance and framework has been reduced from 1,560 in NewLEAF2 to 1,213 in NewLEAF3. It is hoped that by reducing choice and clarifying the guidance and framework, greater agreement can be achieved.

9.2.2 Limitations

Limitations in the methodologies of previous studies have been discussed in previous chapters. This section discusses limitations of the dissertation as a whole.

9.2.2.1 Identification

It seems that issues relating to error identification have not been confidently overcome. The analyst needs to be quite confident with the metalanguage used in the guidance, and confident in applying rules consistently during error identification.

The greatest issue in error identification is acceptability: interpretations of correctness of language appears to be subjective, and there is no comprehensive 'rule book' for analysts to refer to. One analyst may find an utterance acceptable whilst another may identify it as an error. I do not see a way to resolve this issue at the moment and it remains a challenge for inter-rater reliability. However, it is hoped that analysts using the framework will become more familiar with the guidance with use. The guidance seems complex for first-time users, but practice appears to reduce this effect. Dušková (1969;14) states that in her study 'Fortunately, the number of cases in which it was hard to decide whether or not an error had been made....did not exceed 4% of all the errors made'. This seems to be an acceptably small figure: one which was not exceeded in Study 6.

9.2.2.2 Categorisation

Lott (1985:259) states there is difficulty 'in building a system of definitions where the analyst will not periodically have doubts about how to categorise particular errors: the quality of any research must be affected by the researcher's intuitions about [the type of error]'. NewLEAF3 has minimised this issue greatly. However, it is quite possible that future LEAs, using NewLEAF3 on different essay data, will expose more dual categorisation possibilities, but there would be a negligible amount, and this could also be seen as satisfactory.

9.2.2.3 General

Given some historical and current issues with EA and LEA, it is doubtful that any framework and guidance for LEA will be 100% foolproof, and the best that can be hoped for is the minimisation of issues to make the process as accurate and useful as possible.

The framework only employs quantitative analysis. This is an issue as it may not take into account problems with functional aspects of language in learner compositions. Nor does it acknowledge that more advanced learners may attempt and make errors with more complicated language structures (Frodesen 2009 and Doolan and Miller 2012). A qualitative description of learner error would be able to take these points into account, which would be useful for instructional and assessment purposes. Furthermore, a purely quantitative approach is subject to natural limitations. As with all systems of marking and feedback, some qualitative comment is required to address other areas of writing, such as organisation or structure. NewLEAF3 therefore cannot be viewed as a 'one-stop-shop' for feedback on writing, but I do feel that it is quite straightforward to use and will provide a depth of results that teachers and learners will find very useful.

If it had occurred to me to time the LEAs using the different frameworks used in the different studies, it would have enabled discussion of the time efficiency of the various frameworks that have been used in this dissertation.

9.3 Benefits of NewLEAF3

Despite the limitations of the individual studies and the thesis as a whole, the work has produced a framework and guidance that is better than previous attempts at LEA. The results when using NewLEAF3 are more in depth and the framework is easier to use than previous versions and other published frameworks. I believe that I have demonstrated that the framework and guidance are satisfactory when used by someone with some LEA familiarity and the results when using NewLEAF3 will be of great interest and use to teachers and learners.

NewLEAF3 has not yet been tested for inter-rater agreement, but it has massively improved upon previous attempts at LEA and is as workable as it can be when considering historical issues of error identification, separation and categorisation. Better inter-rater agreement scores in terms of categorisation could be achieved by reducing the number of categories, such as [Formality], [Underspecification] and [Verbosity], leaving only [Phrase] error in Section B, but this would reduce the depth of analysis and therefore be of less use to teachers and students. To my mind, an optimum balance between ease of use and detail of results has been achieved.

I would like to repeat here that inter-rater agreement of identification and categorisation is only an issue when more than one analyst is used to compare results of LEA performed on the work of different groups or individuals, or the same groups or individuals when different analysts are used. Intra-rater agreement is a separate issue. NewLEAF3 will be of use to the researcher or teacher when he or she is working as a sole analyst, as long as he or she can be internally systematic in the identification and categorisation of LEs. However, this has not yet been tested and is a possible direction for further research. That said, in Study 5, the six participants that were reused from Study 4 had a higher agreement score in terms of categorisation than the rest of Group A and Group B. This is probably due to the fact that Study 5 was their second LEA experience, which points to the possibility of greater agreement coming with familiarity: a sentiment stated by participants in Study 4.

9.4 Future improvements to NewLEAF3

Perhaps the framework could be further refined to reduce the over-grouping in the most common category: [Phrase errors (misc)]. Many of these could be placed in sub-categories such as [Noun phrase], [Verb phrase] or [Prepositional phrase].

9.5 Future hopes for NEWLEAF and LEA research

In section 1.2 a number of potential users, uses and contexts were identified for use of the tool. In light of the experimental findings, I will now revisit these with comments on the suitability of the tool for these various areas.

The metalanguage used in the framework has been reduced, and with more example sentences and corrections, it is now a more accessible feedback mechanism to students, especially in a one-to-one context, provided that their proficiency level is good enough to understand it (probably B1 and above). This will help to raise awareness of the types and frequency of lexical errors that they are making and will boost autonomy as the learner takes individual steps to eradicate these errors.

This tool would be particularly useful in high-stakes exam preparation classes and EAP classes where there is a higher demand for accuracy and probably less of a priority in more fluency-based courses, such as ESOL in English-speaking countries or general EFL contexts where English is taught as a second or third

language. In countries that have their own varieties of English, the tool could be adapted and used by experts in those varieties so that the learner is not penalised for not conforming to a variant of English that is not relevant to the local context.

The results will be of use for teachers as it will allow them to gather information on the type and frequency of lexical errors that the group is making, and where learning has taken place and where it has not; it can also inform remedial teaching. This should also help to inform teachers of which learning methods have been more successful than others.

The framework could be used to standardise rater grades for lexical resource. Despite there still being issues with inter-rater agreement for lexical error identification and categorisation, the guidance will help those with more of a background in linguistics to mark more similarly than if there were no guidance, especially if inter-rater agreement increases with experience or use of the tool.

The tool could be used at the start of a course, especially a high-stakes ESP course, to diagnose issues with use of lexical items. This would inform syllabus selection before teaching commences.

In the same way, materials writers' work will be informed for the design of general coursebook materials, if they were to be made more aware of the types and frequency of lexical errors that general English learners make.

However, the unresolved issues with inter-rater identification and categorisation agreement makes the tool more appropriate for use by a single rater, rather than a team of raters. Especially, the lack of agreement between raters about what a lexical error is means that its use as a tool in second language research when there is more than one researcher involved might be less reliable than desired. Conversely, it seems logical that the issue of inter-rater categorisation agreement will be reduced if there are fewer categories and more training or experience using the tool.

Further adaptations could be made to the tool to make it more useful in certain contexts. For example, as discussed in Section 5.2, it could be adapted to incorporate an analysis of punctuation, or to analyse pronunciation errors or grammatical ones. However, staying with a focus on lexis, some categories such

as formality, could be removed for lower proficiency levels so that number of errors exposed is not too high, as this may lead to demotivation for the learner. Some teachers may wish to make more of the final stage of error evaluation when deciding on which errors to focus upon; for example, errors of form which can be assessed objectively, e.g. spelling errors rather than more subjective errors, such as those involving social awareness or the tone of particular lexical items. Errors made with core vocabulary (Bell 2012) could be prioritised for remedial teaching, or errors with items that were previously taught could be the focus of a review.

Again, NewLEAF should not be seen or used as the sole feedback mechanism, as it does not focus on what the learner has done well in terms of lexical choices. Nor does it, in its current form, analyse use of punctuation, cohesion, organisation, etc, etc. It should be viewed as a method of giving super-refined feedback on a very specific area of accurate language use.

I will publish the framework in a journal, such as the English Language Teaching Journal, so that it could be shared with practising language teachers so that they will be able to use it not only to investigate the errors that their learners make and share the information with them to improve teaching and learning, but also to share results with the wider teaching community.

An online forum where teachers could share their LEA results would enable comparison of data, and establish similarity of error type and frequency across L1 groups and therefore inform teaching. This would also enable us to more confidently confirm or refute Hemchua and Schmitt's (2006) statement that learners of different L1 groups will make similar types of lexical errors.

I will also publish in academic journals so that it reaches an audience of SLA researchers. In this area, the next steps would be to improve the framework so that the issues mentioned above could be overcome. Then, it would be very informative if the framework could be used to analyse similar-length essay data from similar proficiency level cohorts of different L1 groups to establish similarity of findings. Next, it would be interesting to use the framework on different proficiency levels to further test Llach's (2011) statement that learners of a higher proficiency level make more semantic errors than formal ones and establish the specific types and frequencies of those errors.

9.6 Final thoughts

The idea for this research was inspired by the work of authors such as Hemchua and Schmitt (2006) and by my own experience as an English language teacher and tester. However, during the course of this dissertation, my view of language has evolved from a system that could possibly be classified as containing elements that are right or wrong by different raters, to one that is less black and white, where 'correctness' may vary considerably from person to person and is probably dependent on background and training, as this determines what they feel is appropriate for the genre in which a learner is writing. To some degree, issues with acceptability were expected, but I wanted to establish whether these could be overcome with detailed guidance and better categories. Whilst a satisfactory method for identification to ensure inter-rater agreement has not been created, this research has thoroughly investigated and provided some methods of mitigating some issues with EA so that LEA may meet its potential. I still believe strongly that LEA is a very useful tool to assist in exam marking, for individual teachers to ascertain weaknesses in individual or whole cohort writing and as a feedback tool for learners.

References

- Allerton, D. (1990) 'Linguistically strange word combinations.' In Bridges, M. (ed) *On strangeness* Tübingen: Gunther Narr.
- Al-Khresheh, M. (2015). 'A review study of interlanguage theory'. *International Journal of Applied Linguistics and English Literature*, 4(3) pp. 124-131.
- Al-Khresheh, M. (2016) 'A review study of error analysis theory'. *International Journal of Humanities and Social Science Research*, 2 pp. 49-59.
- Al-Shormani, M. Q. and Al-Sohbani Y. A., (2012) 'Semantic errors committed by Yemeni university learners: classifications and sources.' *International Journal of English Linguistics* 2(6) pp. 120-139.
- Al-Shormani, M. Q. (2014a) 'Lexical choice difficulties: a psycholinguistic study towards a solution'. *Journal of Literature, Languages and Linguistics* 4 pp. 43-53.
- Al-Shormani, M. Q. (2014b) 'Collocability difficulty: a UG-based model for stable acquisition.' *Journal of Literature, Languages and Linguistics*.4 pp. 54-64.
- Astika, G. C. (1993) 'Analytical assessment of foreign students' writing.' *RELC Journal* 24 pp. 61-72.
- Azevedo, R., Taub, M. and Mudrick, N. (2015) 'Think-Aloud Protocol Analysis' in Spector, J.M. (ed) *The Sage encyclopaedia of educational technology* Thousand Oaks: Sage Publications.
- Bardovi-Harlig, K. and Bofman, T. (1989) 'Attainment of syntactic and morphological accuracy by advanced language learners' *Studies in Second Language Acquisition*, 11(1), pp. 17-34.
- Behydt, L. (1987) 'The semantization of vocabulary in foreign language learning.' *System*, 15(1) pp. 55-67.
- Bell, H. 2009. 'The Messy Little Details: A Longitudinal Case Study of the Emerging Lexicon'. In Fitzpatrick, T. and A. Barfield (eds.) *Lexical Processing in Second Language Learners Papers and Perspectives in Honour of Paul Meara*. Bristol: Multilingual Matters
- Bell, H. (2012) 'Core Vocabulary. In C.A. Chapelle (ed.) *The encyclopaedia of applied linguistics*. Oxford: Wiley-Blackwell.
- Bell, R.T. (1974) 'Error analysis a recent pseudoprocedure in applied linguistics.' *International Review of Applied Linguistics*. 25/26 pp. 35-49.
- Benson, C. (2002) 'Transfer/Cross-linguistic influence' *ELT Journal* 56/1
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Longman.

- Bley-Vroman, R. (1983) 'The comparative fallacy in interlanguage studies: the case of systematicity.' *Language Learning* 33 (1) pp. 1-17.
- Bodle, A. (2016) *How new words are born* The Guardian [Online] [Accessed on 10th Jan 2021] <https://www.theguardian.com/media/mind-your-language/2016/feb/04/english-neologisms-new-words>
- Brown, H. D. (2000). *Principles of language learning and teaching* 4th ed., San Francisco State University: Pearson Education Limited.
- Burt, M. K., and Kiparsky, C. (1972). *The gooficon: A repair manual for English*. Rowley MA: Newbury House.
- Burt, M. (1975) 'Error analysis in the adult EFL classroom.' *TESOL Quarterly* 9 (1) pp.53-63.
- Burton, D. and Bartlett, S. (2009) *Key issues for education researchers*. London: SAGE.
- Buyl, A. and Housen, A. (2015) 'Developmental stages in receptive grammar acquisition: a processability theory account'. *Second Language Acquisition* 31(4) pp. 523-550.
- Bygate, M. (1987) *Speaking*. Oxford: Oxford University Press.
- Candler W. J. (1979) 'Errors, error, and text in a multidialect setting.' *ELT Journal*, 33(4) pp. 266–274.
- Carter, R. (2012) *Vocabulary: applied linguistics perspectives*. (2nd ed) London: Routledge.
- Chamot, A. (1979) 'Strategies in the acquisition of English structures by a child bilingual in Spanish and French.' In R. Andersen (ed.) *The acquisition and use of Spanish and English as first and second languages*. Washington D.C.: TESOL.
- Chomsky, N. (1957) 'Syntactic Structures'. The Hague: Mouton.
- Chomsky, N. (1965) 'Aspects of the theory of syntax'. Cambridge, MA: MIT Press.
- Clark, H. and Clark, E. (1977) 'Psychology and language'. *Journal of Child Language* 4(2) pp. 366-382.
- Cohen, L., Manion, L. and Morrison, K. (2018) *Research methods in education*. 7th ed. Hoboken: Taylor and Francis.
- Cook, V. J. (1993) *Linguistics and second language acquisition*. Basingstoke: Macmillan.
- Corder, S. P. (1967) The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5 (1-4) pp. 161-170.
- Corder, S. P. (1971) Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics in Language Teaching*, 9(2) pp. 147-160.

- Corder, S. P. (1973a) *Introducing error analysis*. Middlesex: Penguin.
- Corder, S. P. (1973b) *Introducing applied linguistics*. Middlesex: Penguin.
- Corder, S. P. (1974) *Error analysis*. Oxford: Oxford University Press.
- Corder, S. P. (1981) *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Coyle, Y. and Roca de Larios, J. (2014) Exploring the role played by error correction and models on children's reported noticing and output production in an L2 writing task. *Studies in Second Language Acquisition* 36(3) pp. 451-485.
- Creswell, J.W. and Plano, V.L. (2011) *Designing and conducting mixed methods research*. 2nd ed., London: Sage.
- De Rocha, F.J. (1980) 'On the reliability of error analysis studia.' *Anglica Posnaniensia* 12 pp. 83-90.
- Descombe, M. (2014) *Good research guide: for small-scale research projects*. Maidenhead: McGraw-Hill Education.
- Dilans, G. (2010) 'Corrective feedback and L2 vocabulary development: prompts and recasts in the adult ESL classroom'. *Canadian Modern Language Review* 66(6) pp. 787-816.
- Doolan, S.M. and Miller, D. (2012) 'Generation 1.5 written error patterns: a comparative study.' *Journal of Second Language Writing* 21(1) pp. 1-22.
- Dörnyei, Z. (2007) *Research methods in applied linguistics* Oxford: Oxford University Press.
- Dulay, H. C., and Burt, M. K. (1974). 'Errors and strategies in child second language acquisition.' *TESOL Quarterly*, 8(2) pp. 129-136.
- Dulay, H. Burt, N. and Krashen, S. (1982) *Language two* Oxford: Oxford University Press.
- Dušková, L. (1969) 'On sources of errors in foreign language learning'. *International Review of Applied Linguistics*. 7 pp. 11-36.
- Ellis, R. (1985) 'Sources of variability in interlanguage.' *Applied Linguistics* 6: pp. 118-131.
- Ellis, R. (1994) 'A theory of instructed second language acquisition' In N. Ellis (ed) *Implicit and explicit learning of languages*. San Diego: Academic Press.
- Ellis, R. (1996) *The Study of Second Language Acquisition* Oxford: Oxford University Press
- Ellis, R. (1997) *SLA research and language teaching* Oxford: Oxford University Press.

- Ellis, R. (2003) *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2005) 'Instructed language learning and task-based teaching', in E. Hinkel (ed.) *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum. pp. 713–728.
- Ellis, R. and Barkhuizen (2005) *Analysing Learner Language* Oxford: Oxford University Press
- Ellis, R. (2008) *The study of second language acquisition 2nd Edn* Oxford: Oxford University Press.
- Ellis, R. (2016) 'Focus on form: a critical review.' *Language Teaching Research*, 20(3) pp. 405-428.
- Ellis, R. and Barkhuizen, R. (2005) *Analysing learner language*. Oxford: Oxford University Press.
- Ellis, R., Basturkmen, H., and Loewen, S. (2001) 'Learner uptake in communicative ESL lessons.' *Language Learning*, 51(2) pp. 281-318.
- Engber, C. A. (1995) 'The relationship of lexical proficiency to the quality of EFL compositions.' *Journal of Second Language Writing* 4(2) pp. 139-155.
- Erdogan, V. (2005) 'Contribution of error analysis to foreign language teaching'. *Mersin University Journal of the Faculty of Education*, 1(2) pp. 261-270.
- Ericsson, K. A., and Simon, H. A. (1993) *Protocol analysis: verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.
- Etherton, A.R.B (1977) 'Error analysis: problems and procedures.' *ELT Journal*, 32(1) pp. 67–78.
- Falk, J. (1978). '*Linguistics and Language: A Survey of Basic Concepts and Implications (2nd ed.)*'. New York: John Wiley and Sons.
- Felix, S. (1981) 'The effect of formal instruction on second language acquisition.' *Language Learning* 31(1) pp. 87-112.
- Ferris, D.R. (2002) *Treatment of error in second language writing*. Anne Arbor MI: University of Michigan Press.
- Ferris, D. (2009) *Teaching college writing to diverse student populations*. Anne Arbor, MI: Michigan University Press.
- Foin, A.T. and Lange, E.J. (2007) 'Generation 1.5 writers' success in correcting errors marked on an out-of-class paper.' *The CATESOL Journal*, 19(1) pp. 146-163.
- Foster, P., and Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3) pp. 299-324.
- Fries, C. (1945) *Teaching and learning English as a foreign language*. Ann Arbor: University of Michigan Press.

- Frodesen, J. (2009) The academic writing development of a Generation 1.5 latecomer *In* M. Roberge, M. Siegal, L. Harklau (ed.) *Generation 1.5 in college composition*. New York: Routledge pp. 91-104.
- Frodesen J. and Starna, N. (1999) Distinguishing incipient and functional bilingual writers: assessment and instructional insights gained through second-language writing profiles. *In* L. Harklau, M. Kay, M. Siegal (ed.), *Generation 1.5 meets college composition: issues in the teaching of writing to U.S.-educated learners of ESL*. Mahwah, NJ: Erlbaum, pp. 61-80.
- Gamallo, M. G., del Río, I and González. I. (2015) 'Avalingua: natural language processing for automatic error detection in learner corpora'. *In* Callies, M. (ed.) *Learner Corpora in Language Testing and Assessment*. University of Bremen: John Benjamins Amsterdam B.V. Volume 70.
- Gass, S. and Selinker, L. (1992) Eds. *Language Transfer in Language Learning: Revised edition* Amsterdam: John Benjamins
- Given, L. M. (2008) *The Sage encyclopaedia of qualitative research methods*. London: SAGE.
- George, H.V. (1972) *Common errors in language learning: insights from English*. Rowley, MA: Newby House.
- Ghadessy, M. (1980). 'Implications of error analysis for second/foreign language acquisition.' *International Review of Applied Linguistics in Language Teaching*, 18 pp. 93-104.
- Gordon, T., & Harshbarger, S. (2003) 'The questions of grammar: The place of grammar in L2 teacher education'. *Grammar teaching in teacher education*. pp. 41-53.
- Götz, S. (2015) 'Tense and aspect errors in spoken English: Implications for Language Testing and Assessment'. *In* Callies, M. *Learner Corpora in Language Testing and Assessment*. University of Bremen: John Benjamins Amsterdam B.V. Volume 70.
- Grauberg, W. (1971) *An error analysis in German of first year university students*, Longman: London.
- Granger, S. (1998) *Learner English on computer* London: Addison Wesley Longman.
- Granger, S. (2002) 'A bird's-eye view of computer learner corpus research.' *In* S. Granger, J. Hung and S. Petch-Tyson (ed.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching [Language Learning and Language Teaching 6]*. Amsterdam: John Benjamins. pp. 3-33.
- Granger, S. (2008) Learner corpora. *In* A. Lüdeling (ed.) *Corpus Linguistics. An International Handbook*. Berlin: Walter de Gruyter pp. 259–275.
- Granger, S. and Tyson, S. (1996) Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*. 15(1) pp. 19-29.

- Halliday, M. A. K. (1966) '*Lexis as a linguistic level*'. In C. Bazzell, J.C. Catford, M.A.K. Halliday & R.H. Robin (Eds), In memory of J.R. Firth, London: Longman
- Halliday, M.A.K. (1985). '*Systemic Background*'. In "Systemic Perspectives on Discourse, Vol. 1: Selected Theoretical Papers" from the Ninth International Systemic Workshop, James D. Benson and William S. Greaves (eds). Ablex. Reprinted in Full in Volume 3 in The Collected Works of M.A.K. Halliday. London: Continuum.
- Halliday, M.A.K. (2013) '*Halliday's Introduction to Functional Grammar. 4th ed.*' New York: Routledge
- Hammarberg, B. (1973) *The Insufficiency of Error analysis* In J. Svartvik (ed.) *Errata: papers in error analysis*. Lund, Sweden: CWK Gleerup.
- Hanaoka, O. (2007) 'Output, noticing and learning: an investigation into the role of spontaneous attention to form in a four-stage writing task.' *Language Teaching Research* 11(4) pp. 459-479.
- Hancock, P. (1990) *Is that what you mean?* London: Penguin.
- Harmer, J. (2015) *The practice of English language teaching*. London: Pearson.
- Hendrickson, J. (1978) 'Error correction in foreign language teaching: recent theory, research and practice.' *The Modern Language Journal* 62(8) pp. 387-98.
- Hemchua, S. and Schmitt, N. (2006) *An analysis of lexical errors in the English compositions of Thai learners* Prospect 21(3) pp. 3-25.
- Ho-Peng, L. (1983) 'Using T-unit measures to assess writing proficiency of university ESL students' *RELC Journal*. 14(2) pp. 35-43.
- Housen, A., Kuiken, F. and Vedder, I. (2012) *Dimensions of L2 performance and proficiency: complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing.
- Howarth, P. (1998) 'Phraseology and second language proficiency'. *Applied Linguistics*, 19(1) pp. 22-44.
- Hunston, S., G. Francis, and E. Manning (1997) 'Grammar and vocabulary showing the connections' *ELTJ* 51(3) pp. 208-216.
- Hunston, S. and Francis, J. (2000) *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing.
- Ishikawa, S. (1995) 'Objective measurement of low-proficiency EFL narrative writing.' *Journal of Second Language Writing*, 4(1) pp. 51-70.
- Jain, M. (1974) Error analysis: source, cause and significance In Richards, J.C. (ed.) *Error Analysis*. London: Longman.
- James, C. (1977) Judgements of Error Gravities *ELTJ* 31(2) pp. 116-24.

- James, C. (1998) *Errors in language learning and use: exploring error analysis*. New York: Longman.
- Jenkins, J. (2006) Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL quarterly*, 40(1), 157-181.
- Jiang, W. (2009) *Acquisition of word order in Chinese as a foreign language*. Germany: Mouton de Gruyter.
- Johansson, S. (1973) 'The identification and evaluation of errors in foreign languages: a functional approach.' in Svartnik, J. (ed.) *Errata Papers in Error Analysis* Lund: Sweden CWK Gleerup.
- Kallkvist, M. (1998) 'How different are the results of translation tasks? A study of lexical errors.' In Malmkjaer K. (ed) *Translation and language teaching: Language teaching and translation*. Manchester: St. Jerome Publishing pp. 77-87.
- Kamiya, N. and Nakata, T. (2021) 'Corrective feedback and the development of second language vocabulary.' In Nassaji, H. and Kartchava, E. (ed) *The Cambridge Handbook of Corrective Feedback in Second Language Teaching and Learning*. Cambridge: Cambridge University Press.
- Kasper, G., and Schmidt, R. (1996) 'Developmental issues in interlanguage pragmatics.' *Studies in Second Language Acquisition*. 18(2) pp. 149-169.
- Kellerman, E. (1995) 'Crosslinguistic influence: transfer to nowhere?' *Annual Review of Applied Linguistics*. (15) pp. 125-150.
- Khalil, A. (1985) 'Communicative error evaluations: native-speaker evaluation and interpretation of written errors of Arab EFL learners' *TESOL Quarterly* 19(2) pp. 225-351.
- Krashen, S., and Terrell, T. (1983) *The natural approach*. San Francisco, CA: Alemany Press.
- Krashen, S.D. (1985) *The input hypothesis: issues and implications*. New York: Longman.
- Kreyer, R. (2015) 'The Marburg Corpus of Intermediate Learner English (MILE)' in Callies, M. (ed.) *Learner Corpora in Language Testing and Assessment* (ed.) University of Bremen: John Benjamins Amsterdam B.V. Volume 70.
- Larsen-Freeman, D. and Strom, V. (1977) 'The construction of a second language acquisition index of development.' *Language Learning*, 27(1) pp. 123-134.
- Laufer, B. (1990) 'Why are some words more difficult than others? Some intralexical factors that affect the learning of words.' *International Review of Applied Linguistics*. 28, pp. 293-307.
- Laufer B. (1991) 'Some properties of the foreign language learner's lexicon, as evidenced by lexical confusions.' *International Review of Applied Linguistics*. 29(4) pp. 317-330.

- Laufer, B. (1992) 'How much lexis is necessary for reading comprehension?' In P. Arnaud and H. Bejoint (ed.) *Vocabulary and Applied Linguistics*. London: Macmillan. pp.126-132.
- Lee, S. (2017) Raising EFL learners' awareness of L2 lexical errors and correct usage: a dual coding approach.' *English Teaching*. 72(2) pp. 29-50.
- Leech, G. N. (1981) *Semantics*. 2nd ed., Middlesex: Penguin.
- Leki, I. (1991) 'The preference of ESL students for error correction in college-level writing classes.' *Foreign Language Annals*, 24(3) pp. 180-203.
- Lennon, P. (1991a) 'Error and the very advanced learner' *International Review of Applied Linguistics* 29(1) pp. 31-43.
- Lennon, P. (1991b) 'Error: some problems of definition, identification, and distinction.' *Applied Linguistics* 12(2) pp. 180-196.
- Lewis, M. (1993) *The lexical approach*. Hove: Language Teaching Publications.
- Lewis, M. (2000) 'Language in the lexical approach.' In Lewis, M. (ed.) *Teaching Collocations*. Hove: Language Teaching Publications.
- Lewis, M. (2000) *Teaching collocation: further developments in the lexical approach*. Boston: Heinle and Heinle.
- Li, S. (2010) 'The effectiveness of corrective feedback in SLA: a meta-analysis.' *Language Learning*, 60(2) pp. 309-365.
- Li., S. (2014) 'Oral corrective feedback'. *ELT Journal*. 68(2) pp.196-198.
- Llach, M. P. A. (2005a) 'The relationship of lexical error and their types to the quality of ESL composition: An empirical study.' *Porta Linguarum*, 3 pp. 45-57.
- Llach, M. P. A. (2007a) 'Lexical errors in young EFL learners: How do they relate to proficiency measures?' *Interlinguistica*, 17, pp. 63-73.
- Llach, M. P. A. (2007b) 'Lexical errors as writing quality predictors.' *Studia Linguistica*. 61(1) pp. 1-19.
- Llach, M. (2011) *Lexical errors and language accuracy in foreign language writing*. Bristol: Multi-Lingual Matters.
- LoCoco, V. (1976) 'A comparison of three methods for the collection of L2 data: free composition, translation and picture description.' *Working Papers on Bilingualism* 8 pp. 58-86.
- Long, M. H. (1996) 'The role of linguistic environment in second language acquisition.' In Ritchie, W. C. and T. K. Bhatia (ed.) *Handbook of Second Language Acquisition*. San Diego: Academic Press. pp. 413-468.
- Long, M., and Robinson, P. (1998) 'Focus on form Theory, research, and practice.' In M. Swain, C. Doughty, and J. Williams (ed.), *Focus on form in classroom*

- second language acquisition*. Cambridge: Cambridge University Press. 15 pp. 15-41.
- Long, M. (2014) *Second language acquisition and task-based language teaching* Chichester: Wiley-Blackwell.
- Long, M. and Sato, C. (1984) 'Methodological issues in interlanguage studies: an interactionist perspective.' In A. Davis, C. Cramer, and A. Howatt (ed.) *Interlanguage*. Edinburgh: Edinburgh University Press.
- Lott, D. (1985) 'Analysing and counteracting interference errors.' *ELTJ* 37(3) pp. 256-261.
- Lüdeling, A., Walter, M., Kroymann, E. and Adolphs, P. (2005) 'Multi-level error annotation in learner corpora.' *Proceedings of the Corpus Linguistics Conference Series* 1(1): 105–115.
- Ludwig, J. (1982) Native-speaker judgements of second language learners' efforts at communication: a review. *The Modern Language Journal* 66(3) pp. 274-283.
- Lyons, J. (1968) *An Introduction to Theoretical Linguistics* Cambridge: Cambridge University Press.
- Lyster, R. and L. Ranta. (1997) 'Corrective feedback and learner uptake'. *Studies in Second Language Acquisition* 19(1) pp. 37–66.
- Mackey, A. (2013) *Input, interaction and corrective feedback in L2 learning*. Oxford: Oxford University Press.
- Mackey, A.J., Park, H.I. and Tagarelli, K. (2016) 'Error, feedback, and repair: variations and learning outcomes.' In Hall, G. (ed.) *Routledge Handbook of English Language Teaching*. Routledge, London.
- Manchester Metropolitan University (2016) *Research Regulations and Guidance* [Online] [Last accessed 10th January 2016] <https://www.mmu.ac.uk/study/postgraduate/research-study/phd/research-regulations-and-guidance>
- Martin, M. (1984) 'Advanced vocabulary teaching: the problems of synonyms.' *The Modern Language Journal* 68(2) pp. 130-137.
- Meara, P. (1980) 'Vocabulary acquisition: a neglected aspect of language learning.' *Language Teaching and Linguistics Abstracts* 13(4) pp. 221-246.
- Meara, P. (1984) 'The study of lexis in interlanguage.' In Davis, A. Cramer C. and Howatt A. (ed.) *Interlanguage*. Edinburgh: Edinburgh University Press.
- Meara, P. and English, F. (1987) *Lexical errors and learners' dictionaries*. ERIC Reports-Research/Technical. [Online] [Last Accessed 01.02.16] 143.<https://eric.ed.gov/?q=Lexical+errors+and+learners%E2%80%99+dictionaries&id=ED290322>.
- Nemser, W. (1971) 'Approximate systems of foreign language learners.' *International Review of Applied Linguistics*. 9(2) pp. 115-23.

- Norrish, J. (1983) *Language learners and their errors*. London: Macmillan.
- Odlin, T. (1989) *Language Transfer* Cambridge: Cambridge University Press
- Odlin, T. (2022) *Explorations of Language Transfer*. Bristol, Blue Ridge Summit: Multilingual Matters.
- Papangkorn, P. (2015) 'SSRUIC students' attitude and preference toward error corrections.' *Procedia - Social and Behavioral Sciences*. 197 pp. 1841-1846.
- Pavesi, M. (1986) 'Markedness, discursial modes and relative clause formation in formal and informal contexts.' *Studies in Second Language Acquisition* 8(1) pp. 38-55.
- Pica, T. (1984) Methods of Morpheme Quantification: their effect on the interpretation of second language data *Studies in Second Language Acquisition* 6: 69-78
- Picot, A. (2017) 'Lessons in lexical error analysis. revisiting Hemchua and Schmitt (2006): an analysis of the lexical errors in the compositions of Greek learners.' *International Journal of English Language Teaching* 5(4) pp. 40-59.
- Politzer, R. L., and Ramirez, A. G. (1973). 'An error analysis of the spoken English of Mexican-American pupils in a bilingual school and a monolingual school.' *Language Learning*. 23(1) pp. 39-61.
- Porte, G. (2012) *Replication research in applied linguistics* (ed.) Cambridge: Cambridge University Press.
- Reid, J.M. (1997) 'Which non-native speaker? Differences between international students and U.S. resident (language minority) students.' *New Directions for Teaching and Learning*, 70 pp. 17-27.
- Richards, J. (1971) 'A non-contrastive approach to error analysis.' *English Language Teaching* 25() pp. 204-219.
- Richards, J. C. (1976). *Error analysis: perspectives on second language acquisition*. Longman: London.
- Richards, J. C. and Rogers, T (2001) *Approaches and Methods in Language Teaching* Cambridge: Cambridge University Press.
- Richards, J. C. (2002) 'Addressing the grammar gap in task work.' In Richards, J. C. and Renandya, W. (ed.) *Methodology in Language Teaching: An Anthology of Current Practice*. Cambridge: Cambridge University Press.
- Richards, J.C. (2015) *Key Issues in Language Teaching*. Cambridge: Cambridge University Press.
- Salem, I. (2007) 'The Lexico-grammatical continuum viewed through student error.' *ELTJ*. 61(3) pp. 211-219.
- Santos, T. (1988) 'Professors' reactions to the writing of non-native-speaker students.' *TESOL Quarterly*. 22(1) pp. 69-90.

- Schachter, J. (1974) 'An error in error analysis.' *Language Learning*. 24(2) pp. 205-214.
- Schachter, J., and Celce-Murcia, M. (1977) 'Some reservations concerning error analysis.' *TESOL Quarterly*. 11(4) pp. 441-451.
- Schmidt, R. (1990) 'The role of consciousness in second language learning.' *Applied Linguistics*. 11(2) pp.129-158.
- Schmitt, N. (ed.). (2004) *Formulaic Sequences: Acquisition, Processing, and Use* John Benjamins Press
- Schmidt, R. (2001) 'Attention' In P. Robinson (ed.) *Cognition and second language acquisition*. Cambridge: Cambridge University Press. pp. 3–32.
- Schubert, S. (2002) "'Are you wanting a cup of coffee?'" Overuse of the progressive aspect in Indian English [Online] [Last accessed 15 Jan 2021] <http://kops.uni-konstanz.de/handle/123456789/3793;jsessionid=F71131A203CA2934CC349C6C58CC744C>
- Schuman, J. H., and Stenson, N. (1974) *Induced errors. new frontiers in second-language learning*. Rowley, MA: Newbury House.
- Selinker, L. (1972) Interlanguage. *International Review Of Applied Linguistics*, 10(3) pp. 209-231.
- Selinker, L. (1992) *Rediscovering interlanguage*. London: Longman.
- Shaalani, K., Magdy, M., and Fahmy, A. (2015) 'Analysis and feedback of erroneous Arabic verbs' *Natural Language Engineering*. 21(2) pp. 271–323.
- Shaw, P. and Liu, E. (1998) 'What develops in the development of second-language writing?' *Applied Linguistics*. 19 pp. 225-254.
- Sheen, Y. (2010) 'Differential effects of oral and written corrective feedback in the ESL classroom'. *Studies in second language acquisition*. 32(2). pp. 203-234.
- Sinclair, J. (1991) *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Spillner, B. (1991) *Comprehensive Bibliography of Error Analysis*. Amsterdam: John Benjamins.
- Spolsky, B. (1979) 'Contrastive analysis, error analysis, interlanguage, and other useful fads'. *The Modern Language Journal*. 63(5/6) pp. 250-257.
- Storch, N. and Wigglesworth, G. (2010) 'Learners' procession, uptake and retention of corrective feedback on writing.' *Studies in Second Language Acquisition*. 32(2) pp. 303-334.

- Swain, M. (2005) 'The output hypothesis: theory and research.' In Hinkel, E. (ed.) *Handbook of research in second language teaching and learning*. New York: Routledge. pp. 495-508.
- Swain, M and Lapkin, S. (1995) 'Problems in output and the cognitive processes they generate: a step towards second language learning.' *Applied Linguistics*. 16(3) pp. 371-391.
- Swan, M. and Smith, B. (2001) *Learner English: a teacher's guide to interference and other problems* (ed.) Cambridge: Cambridge University Press.
- Taylor, G. (1986) 'Errors and explanations.' *Applied Linguistics*. 7(2) pp. 144-166.
- Teh, G. C. (1993) 'Error analysis and correction of written work in the classroom.' *The English Teacher* 22 pp. 1-15.
- Thomas, J. (1984) 'Cross-cultural discourse as unequal encounter.' *Applied Linguistics*. 5(3) pp. 226-35.
- Thornbury, S. (2019) Learning language in chunks. Part of the Cambridge Papers in ELT series. [Online] [Last accessed 15th Jan 2022] www.cambridge.org/cambridge-papers-elt
- Tomiyana, M. (1980) Grammatical errors and communication breakdown *TESOL Quarterly* 14: pp. 71-9.
- Tono, Y., Kaneko, T., Isahara, H., Saiga, T., Izumi, E., Narita, M. and Kaneko, E. (2001) 'The standard speaking test (SST) corpus: a 1million-word spoken corpus of Japanese learners of English and its implications for L2 lexicography.' In S. Lee (ed) *ASIALEX (2001) Proceedings: Asian Bilingualism and the Dictionary. The Second Asialex International Congress*, 8-10 August 2001, Yonsei University, Korea.
- Truscott, J. (1996) 'The case against grammar correction in L2 writing classes.' *Language Learning*. 46(2) pp. 327-369.
- Tschichold, C. (2003) 'Lexically driven error detection and correction.' *CALICO Journal* 20(3) 549-59.
- University of Portland Writing in APA Style [online] [Last accessed 6th April 2023] <http://teaching.up.edu/edresearch/pages/writing-in-apa-style.html#:~:text=Do%20not%20use%20rhetorical%20questions,Avoid%20ad%20homonym%20comments>.
- Van Els, T., T. Bongaerts, G. Extra, C. Van Os, and A. Janssen-Van Dieten (1984) *Applied Linguistics and the Learning and Teaching of Foreign Languages*. London: Edward Arnold.
- Widdowson, H.G. (1972) 'The Teaching of English as Communication'. *ELTJ*, 27(1) pp. 15-19.
- Wilkins, D. (1972) *Linguistics in Language Teaching*. Cambridge, MA: MIT Press.
- Willis, J. (1996) *A Framework for Task Based Learning*. Harlow: Longman.

- Wray, A. (2008) *Formulaic language: pushing the boundaries*. Oxford: Oxford University Press.
- Yang, X., and Xu, H. (2001). *Errors of creativity*. New York: University Press of America.
- Zimmerman, R. (1986a) 'Classification and distribution of lexical errors in the written work of German learners of English.' *Papers and studies in contrastive linguistics* 21 pp. 31-40.
- Zimmermann, R. (1986b) 'Semantics and lexical error analysis.' *EnglischAmerikanischeStudien*, 2(86) pp. 294-305.
- Zimmerman, R. (1987) 'Form-oriented and content-oriented lexical errors in L2 learners. *International Review of Applied Linguistics* 25(1) pp. 55-67.
- Zughoul, M. (1991). 'Lexical choice: towards writing problematic word lists.' *International Review of Applied Linguistics*. 29(1) pp. 45-60.

Appendices

Appendix 3.1 - Consent form

Research Participant CONSENT FORM		
<p>Thank you for allowing me to use your essay in my research.</p> <p>I would like to analyse your writing to establish improvements in English Language teaching and learning. Your anonymity will be protected and you will not be identifiable from the published research.</p> <p>Name of Researcher: Anthony Picot</p> <ol style="list-style-type: none">1. I confirm that I have read and understand the information above for the above study. I have had the opportunity to consider the information, ask questions and have had them answered satisfactorily.2. I understand that my participation is voluntary and that I am free to leave at any time without giving any reason.3. I understand that data collected during this study will be processed in accordance with data protection law and may be used outside the European Union and published in an academic journal.4. I agree for any artefacts I create during participation to remain in the possession of the researcher. Identifiable artefacts will not be used in research outputs.5. I give permission for a fully anonymised version of the data I provide to be deposited in an Open Access repository so that it can be used for future research and learning.6. I agree to take part in the above study.		
Name of Participant	Date	Signature
Name of Researcher	Date	Signature

Appendix 5.1 Instructions for analysts (NewLEAF1)

Collection of scripts

It would be preferable to have all compositions in electronic form (i.e., in MS Word). This will facilitate either in-text highlighting and labelling with the 'Track changes' comments feature, or copying and pasting errors with their co-text into the framework chart so that they can be isolated and then presented to the learners.

A) Identification

An error is defined as 'a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's native-speaker counterparts' (Lennon 1991:182), unless the error is a norm of the variety of English being used (George 1972).

- 1) Following Corder (1973), first identify and highlight all errors in the writing. This stage is required before analysis can begin.
- 2) Do not attempt to differentiate between error, mistake and slip. This would be subjective. Also, see B3 below.
- 3) To separate out grammatical errors, the rules below from Hemchua and Schmitt (2006) should be followed. However, it must be stated that the guidance here is not completely clear:
 - I. Errors in fixed phrases (e.g., '*What's *a matter?*') are collocational and therefore lexical.
 - II. Outside fixed phrases, errors with articles (e.g., omission, overinclusion or misselection) are considered grammatical and should be excluded from the analysis.
 - III. 'Clause errors' (e.g., '*It's not difficult *for getting to a hospital: *While waiting, my hamburger went cold; *The man who he has the car is here*') are considered grammatical errors.
 - IV. 'Sentence errors' are considered grammatical (e.g., '*I didn't think *how kind they were*'). These might include errors in countability, tense, redundancy and subject/verb agreement.
 - V. Only derivational affix errors (e.g., '*He is kind and *considerable*') are included in the lexical category, not plurality, genitive, third person

singular, comparative nor superlative. I.e., errors in derivational affixes are counted, while errors in inflectional affixes are not.

B) Counting

- 1) Erroneous words and collocational phrase errors are included in the count, each counted separately.
- 2) Multiple lexical errors in a sentence are counted separately unless the whole sentence requires rewriting (a [Paraphrase] error) or when a sentence lacks coherence. In the case of the latter, count errors as one per sentence. For example, **all people there aren't in the globe next a few years?* There are three, based on the three phrases (noun phrase: *all people there*, verb phrase: *aren't in the globe* and prepositional phrase *next a few years*).
- 3) Exact duplicates of errors in the same paper should all be included.

C) Categorisation

- 1) If performing LEA for an individual student, paste the specific error with as much co-text as is required to contextualise the error into the corresponding cell in the framework. Alternatively, if you are collecting quantitative information on groups of students, tally the numbers of errors in each cell. This will give totals of types of errors for the group.
- 2) Where dual categorisation is possible, allocate errors to the more specific category. For example, **In the other hand* is both a [Phrase] and a [Preposition] error. The more specific description of this error would be a [Preposition] error. If doubt remains, errors should be allocated to the category which is more serious in terms of communication breakdown towards the upper end of the section.

The categories and example sentences

Section A

A1 Suffix type. Correct root, but incorrect suffix (e.g., *They were very consider*able* <considerate).

A2 Prefix type. Correct root, but incorrect prefix (e.g., *It's quite *inimportant*. <unimportant).

A3 Spelling. Meaning is clear (e.g., *I need to *safe some money*. <save). Words with unclear meaning should be categorised under B1.

A4 Two words should be one (e.g., **Every one will pay* <everyone). Moot points should be omitted from the analysis.

A5 One word should be two. (e.g., *Where is my *tooth brush?* <toothbrush). Moot points should be omitted from the analysis.

Section B

B1 Coherence. Words do not exist in English, or meaning of whole phrase is unclear (e.g., *Where is those that are?<??*).

B2 Wrong word (field error). A 'totally wrong word' or field error category. This category is for when the wrong word and target words exist in the L2 but are not of the same lexical set or meaning area. Could be a 'false friend' (e.g., *I feel *serious* <stressed). Compare with B3.

B3 Wrong word (sense relations error). Errors which show problems discriminating between co-hyponym, antonym, synonym, superonym, hyponym and concepts of meronymy. Errors in this category would be words in the same lexical set or meaning area as the intended word. Compare with B2 (e.g., *We have modern *equipment in our house*. <appliances).

B4 Preposition. Include all preposition errors: the preposition error as part of a collocational phrase (e.g., **In the other hand* <On the other hand) and errors with prepositions of time and place, e.g., *I'll see you *at Saturday* <I'll see you on Saturday).

B5 Cohesion (lexical substitution, pronoun use, conjunctions). Also known as discourse errors. Could be words or phrases, i.e., as lexical substitution errors when referring to the same subject (e.g., Dr Smith/he/the doctor), incorrect pronoun choice (e.g., 'He' for 'she', etc) and wrong choice of conjunction (e.g., 'however' instead of 'and').

B6 Paraphrase. These could simply be infelicitous or non-expert like phrases. Allocate to this category when there is more than one error in a phrase or for problematic phrases which require re-writing. They could simply be grammatical but odd-sounding or non-expert-like. Meaning is clear (e.g., **the number of people is really great*. <There is a large number of people). These phrases could be mainly erroneous due to [Omission], [Over-inclusion], [Misselection] and [Misorder], etc.

B7 Statistically weighted preference. Collocational errors relating to numbers or amounts (e.g., *The Army suffered *big losses* <heavy).

B8 Arbitrary combinations and irreversible binomials errors. Collocational errors relating to fixed phrases or pairs of words (e.g., **hike-hitch* <hitch-hike).

B9 Connotative Meaning. Lexical choice error. The meaning and grammaticality are clear and correct, but also odd (e.g., *He's quite *notorious for the charity work he's done* <famous).

B10 Verbosity/ unnecessary words/ repetition. When too many words are used or repeated unnecessarily. Grammatically correct phrases (e.g., *He bought an apple and* he bought a banana*. <He bought an apple and a banana).

B11 Stylistic underspecification /missing words. Include phrases/sentences that require more detail to clarify meaning or improve expression (e.g., *Although *cars in the country are lower...* <Although there are lower car numbers in the country....).

B12 Formality. For words or phrases that are grammatically correct, but seem overly formal or informal for the genre in use (e.g., **Kids are not allowed in the lounge after 7pm*. <children). (E.g., *I *informed my girlfriend via the medium of the telephone*. <I told my girlfriend on the phone).

B13 Miscellaneous. For any errors that do not fit into the categories above.

Examples of errors taken from Zimmermann (1986), Hemchua and Schmitt (2006) and Picot (2017).

Appendix 6.1 The new guidance and framework (NewLEAF2) Lexical error analysis - instructions for analysts

Summary: There are two main tasks:

1. Identify all lexical errors in student writing.
2. Categorise those lexical errors using a new framework.

Task 1) Identification of Errors.

Underline all the lexical errors. (A lexical error is a lexical linguistic form or combination of forms which would not be produced by a highly skilled English user).

Follow the rules below to identify and exclude the following grammatical errors from the analysis:

1. Article errors (e.g. a/an/the/no article), except in fixed phrases, such as 'What's **a matter?*' Here, the error is collocational and is therefore a lexical phrase error.
2. Inflectional affixes (e.g. work/worked/working or dog/dogs). Do include derivational affix errors (e.g. He is very *consider*able*).
3. Errors with infinitive 'to' (e.g. **Live without love is not *live'*)
4. Gerund errors (e.g. *I like ski*.*)
5. Genitive errors including errors used to show possession. (e.g. *It's Tom* bike, It is *the bike of Tom.*)
6. Comparative and superlative affixes (e.g. *It is *expensiver, It is the *most big*).
7. Clause errors (e.g. *It's not difficult *for getting to a hospital and *While waiting, my hamburger went cold*).
8. Errors in relative clauses including incorrect relative pronoun and referential errors. (e.g. *The man *what lives next door is a doctor and?*)
9. Errors in countability (*Ten items or *less.*)
10. Errors in tense (*Yesterday I *go to school.*)

Separating Lexical Errors

Multiple lexical errors in a phrase are counted separately. However, when a sentence lacks coherence, and it is difficult to separate errors, count them as one per phrase. For example, in the sentence '**all people there aren't in the globe next a few years*', there are three, based on the three phrases (noun, verb and prepositional phrase).

Task 2) Categorise the lexical errors

- Refer to the framework below to categorise lexical errors and decide which type of lexical errors they are according to the codes in the corresponding framework cells. Record the tally marks in the spaces provided.
- Avoid speculation of what caused the error. Simply focus on the actual error. You may consider causality later.
- If the meaning of the error is unclear, consult the full original composition. Then create a plausible reconstruction of the error to help you understand the intended meaning of the error. To ensure consistency of results, categorise the error based not on this reconstruction, but on the error made. If the meaning of the error still cannot be understood, allocate to the coherence category.

NewLEAF2

- Omission = omitted items that are required
- Over-inclusion = extra items that should not be there
- Mis-selection = the wrong items have been chosen
- Mis-order = items are in the wrong order
- Blend = two **correct (not incorrect)** items have been incorrectly combined in some way

Error type	Description/advice	Examples	O-MISSION	OVERIN-CLUSION	MISSELEC-TION	MIS-ORDER	BLEND
Section A Form							
A1 Suffix	Correct root, but incorrect suffix. Do not include verb tense errors (He was go) subject/verb agreement errors (He go) or errors with plurals	<i>They were very consider*able</i> <considerate> <i>The data were analys*t.</i> <analysed>	1	2	3	N/A	4
A2 Prefix	Correct root, but incorrect/missing prefix.	<i>It's quite *inimportant.</i> <unimportant>.	5	6	7	N/A	8
A3 Spelling	Meaning is clear. Words exist in English. Words with unclear meaning should be categorised as B1 .	<i>I need to *safe some money.</i> <save> <i>Put the *folwers in water.</i> <flowers>	9	10	11	12	13
A4 Two words should be one	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-law or ice cream/ ice-cream). Does not include an extra incorrect word (see Section B2).	<i>*Every one will pay.</i> <everyone> <i>I found it *on line.</i> <online>	N/A	N/A	N/A	N/A	14

A5 One word should be two.	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-law or ice cream/ ice-cream). Does not include missing word (see Section B2).	<i>Where is the *paperbin? <paper bin>.</i> <i>Put it on the *dinnertable. <dinnertable></i>	N/A	N/A	N/A	N/A	15
Section B Meaning			O- MISSION	OVERIN- CLUSION	MISSELEC- TION	MIS- ORDER	BLEND
B1 Coherence	Words do not exist in English or meaning of whole phrase is unclear. Only select this category if you have re-read and still cannot be sure of the meaning of the word or phrase.	<i>*Where is those that are? <?></i> <i>I am *lipsh <?></i>	16				
Errors with Single Words (if clearly a misspelling, allocate to A3, e.g. <i>I need to *safe some money. <save></i>)							
B2 Single Word (miscellaneous)	An error with a single word (error could be related or unrelated to correct word). Word exists in English. Does not fit into other B2 categories below.		17	18	19	N/A	20
B2a Conjunction	Include all conjunctions or transition signals here, including multiword items, such as 'in spite of this'.	<i>It was hot *and I took off my coat. <so></i> <i>In *despite of the rain, we went swimming. <spite></i>	21	22	23	N/A	24
B2B noun	Include compound nouns, but not multi-word units/collocations/fixed expressions.	<i>It was hot so I took off my *hand. <coat?></i> <i>We visited the *art museum. <art gallery></i>	25	26	27	N/A	28

B2C adjective		<i>I feel *serious.</i> <stressed> <i>The views were *handsome.</i> <beautiful>	29	30	31	N/A	32
B2D Adverb/Ad verbial Phrase		<i>He sings *deliciously.</i> <i>They are growing *quick.</i> <quickly>	N/A	33	34	N/A	35
B2E Preposition	Prepositions of time and place and dependent propositions, not errors with infinitive 'to'. Include incorrect phrasal verb errors with particles/prepositions here.	<i>I woke up *on 6am.</i> <at> <i>See you *in the corner.</i> <at/on> <i>I was suspicious *about his behaviour.</i> <of>	36	37	38	N/A	39
B2F Pronoun		<i>My mother, *she is ill.</i>	40	41	42	43	N/A
B2G Verb	Include incorrect phrasal verb errors with roots verbs here.	<i>Do you *get the time, please?</i> <have> <i>He *put on the table.</i> <laid>	44	45	46	47	48
Errors that span across more than one word			O- MISSION	OVERIN- CLUSION	MISSELEC- -TION	MIS- ORDER	BLEND
B3 phrase Error (miscellaneous)	Meaning is clear, but phrase is nonexpert-like. There may be more than one error in the phrase (multi-word units). Use for whole problematic phrases which require re-writing.	<i>* The number of people is really great.</i> <There is a large number of people>. <i>*Next a few years.</i> <In the next few years..>	N/A	N/A	49	50	51

B3A Misordering of words in fixed phrases	Collocational errors relating to order of pairs of words or groups of three.	<p><i>*hike-hitch</i> <hitch-hike></p> <p><i>*fro and to</i> <to and fro></p> <p><i>*Dark, tall and handsome</i> <Tall, dark and handsome></p>	52
B3B Verbosity	Too many words are used, or points are repeated unnecessarily. Categorise single extra word in B2 overinclusion above.	<p><i>He bought an apple and* he bought a banana.</i> <He bought an apple and a banana>.</p> <p><i>I like cake. *I like cake because it is sweet.</i> <I like cake because it is sweet></p>	53
B3C Underspecification	Phrases/sentences that require more detail to improve expression. Categorise single missing word in B2 omission above. If meaning is unclear, categorise as B1.	<p><i>Although *cars in the country are lower...</i> <Although there are lower car numbers in the country....>).</p> <p><i>We cut *because tree high.</i> <We cut the trees because they were getting too high.></p>	54
B3D Formality	Phrases that are correct but seem overly formal or informal for the genre. If there is a formality error with a single word, allocate to word class section B2 above.	<p><i>I *informed my girlfriend via the medium of the telephone.</i> <I told my girlfriend on the phone.></p> <p><i>*I am of the opinion that milk stocks in the refrigerator are starting to ebb.</i> <We need more milk.></p>	55

B4 I can't categorise this error using the categories above	Only use this category if the error cannot fit into any category above.	?	56
---	---	---	----

Appendix 6.2 Letter of ethical approval



15/01/2019

Project Title: Lexical Error Analysis of International Students' Compositions

EthOS Reference Number: 0906

Ethical Opinion

Dear Anthony Picot,

The above application was reviewed by the Arts and Humanities Research Ethics and Governance Committee and, on the 15/01/2019, was given a favourable ethical opinion. The approval is in place until 31/07/2020 .

Conditions of favourable ethical opinion

Application Documents

Document Type	File Name	Date	Version
Additional Documentation	MMU-Application-for-Ethical-Approval-September-2015	03/06/2018	1
Additional Documentation	MMU-Research-Insurance-Checklist-v1-0-19-Sept-2016	03/07/2018	1
Additional Documentation	risk assessment_07032018_135302	03/07/2018	1
Additional Documentation	Interview questions for staff	03/07/2018	1
Additional Documentation	Interview Questions for students	03/07/2018	1
Project Proposal	MMU-Application-for-Ethical-Approval-September-2015	18/07/2018	1
Consent Form	Consent form for students	18/07/2018	1
Information Sheet	Partipant-Information-Sheet for staff	18/07/2018	1
Consent Form	Consent form for Staff	04/10/2018	2
Additional Documentation	Consent form for Staff	13/11/2018	2
Additional Documentation	Consent form for students	27/12/2018	3
Additional Documentation	Partipant-Information-Sheet for staff	27/12/2018	3
Additional Documentation	Partipant-Information-Sheet for students	27/12/2018	3

The Arts and Humanities Research Ethics and Governance Committee favourable ethical opinion is granted with the following conditions

Adherence to Manchester Metropolitan University's Policies and procedures

This ethical approval is conditional on adherence to Manchester Metropolitan University's Policies, Procedures, guidance and Standard Operating procedures. These can be found on the Manchester Metropolitan University Research Ethics and Governance webpages.

Amendments

If you wish to make a change to this approved application, you will be required to submit an amendment. Please visit the Manchester Metropolitan University Research Ethics and Governance webpages or contact your Faculty research officer for advice around how to do this.

We wish you every success with your project.

Art and Humanities Research Ethics and Governance Committee

Appendix 6.3 Participant information sheet for teachers 27/12/18 Version 2

Study Title Lexical Error Analysis (looking at the vocabulary errors in students' work).

I would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or would like more information. Take time to decide whether or not to take part.

What is the purpose of the study? I'm researching the number and type of vocabulary errors in students' writing. This is part of my PhD.

Why have I been invited? I have devised a new framework for analysing lexical errors in students writing and I would like to hear the opinion of experienced English Language Teachers on how easy it is to use.

Do I have to take part? It is up to you to decide. I will describe the study and go through the information sheet, which I will give to you. I will then ask you to sign a consent form to show you agreed to take part. I will not use your name in the research and you will not be identifiable. You are under no obligation to participate, and I will not mention the names of the participants to the Director of the Language Centre, nor the Pre-sessional Course Co-ordinator.

You are free to withdraw at any time, without giving a reason. This will not affect you in any way.

What will happen to me if I take part? Nothing, but I will be very grateful.

Expenses and payments? There is no payment with this project.

What will I have to do? You only need to use the framework to analyse five short pieces of writing and agree to be interviewed with the questions below. Then sign a consent form to show that you have agreed to do this.

What are the possible disadvantages and risks of taking part? There are none.

What are the possible benefits of taking part? You may learn, through experience, about how to analyse lexical errors systematically. The information I take from the study will help teachers to understand the types of errors that students make in their writing and may help to improve teaching and learning.

What if there is a problem? If you have a concern about any aspect of this study, you should ask to speak to the researcher who will do his best to answer your questions. You can contact me on a.picot@mmu.ac.uk or phone me on 0161 247 6183 if you have a problem or complaint with the research study. Alternatively, you could contact my Head of Department, Dr Derek Bousfield on d.bousfield@mmu.ac.uk or phone him on 0161 247 3620. Alternatively, if you wish to complain, you can find the University Complaints Procedure, here

<https://www2.mmu.ac.uk/student-case-management/guidance-for-students/student-complaints-procedure/> My supervisor's name is Dr Huw Bell. He can be contacted on h.bell@mmu.ac.uk or 0161 247 6184. If you do not wish to discuss your complaint with the researcher, you should contact the supervisor in the first instance and then the College Research and Innovation (R&I) Manager.

There are no compensation/insurance/indemnity schemes in place in the event of a complaint.

Will my taking part in the study be kept confidential? Your confidentiality will be safeguarded during and after the study by keeping your recording on a password protected drive at all times. After the work has been completed, your name will be removed from the recording. (The recording will be deleted after it has been transcribed.) These methods follow the new General Data Protection Regulations (<https://gdpr-info.eu/chapter-3/>). The recordings will be anonymised and kept confidential by the researcher (A. Picot) on a password protected drive. Individual participant transcripts will be anonymous and given a research code, known only to the researcher. An encrypted master list identifying participants to the research codes data will be held on a password protected computer accessed only by the researcher. Electronic data will be stored on a password protected computer known only by researcher. The data will only be seen and used by myself and only used for my PhD. No further RGEC approval will be sought. The data will be kept for 2 years and then disposed of securely. All information which is collected about you during the course of the research will be kept strictly confidential, and any information about you which leaves the university will have your name removed so that you cannot be recognised

What will happen if I don't carry on with the study? If you withdraw from the study we will destroy all your identifiable documents, but we will need to use the data collected up to your withdrawal.

What will happen to the results of the research study? I hope to publish them with my PhD. Results will be available to you on request. You will not be identified in any report/publication.

Who is organising or sponsoring the research? Funded by Manchester Metropolitan University.

Further information and contact details: If you have any further questions, please do not hesitate to contact me. A.picot@mmu.ac.uk. Many thanks for your participation!

Appendix 6.4 Consent form for students Version 1

Please initial all boxes

I confirm that I have read and understand the information sheet dated 22/6/18 (version 1) for the above study. I have had the opportunity to think about the information, ask questions and hear satisfactory answers.

I understand that I can volunteer for this and I can stop taking part at any time without giving any reason

I have received a written description of the project.

I understand that any writing used will not have my name on it, and that my grades and feedback will not be used or considered in the study in any way.

I agree to the use of my written work.

I understand that parts of my writing without my name may be included in journals or books from this research.

I understand what I need to do as a volunteer in this project.

I understand that I will receive no money for taking part in this project.

I agree to take part in the above study.

_____	_____	_____
Name of Participant	Date	Signature

_____	_____	_____
Name of Person	Date	Signature

taking consent.

Appendix 6.5 Consent form for teachers Version 3



**Manchester
Metropolitan**
University

Project: Lexical Error Analysis of International Students' Compositions

Consent form for Staff

Please initial all
boxes

I confirm that I have read and understand the information sheet dated 27/12/18 Version 2 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.

I give consent for the interviews to be recorded.

I agree to use the provided framework to analyse student work.

I agree to take part in the above study.

Name of Participant

Date

Signature

Name of Person
taking consent.

Date

Signature

Many thanks from the researcher, Anthony Picot

Appendix 6.6 Student's essay

To get a perfect academic performance is a basic quality that every student wants. So, students must develop various skills to achieve the goal. There are many skills to help students to study. Some people think the Reading skill is the most important compared with others. However, others hold the opposite view. Let me discuss the topic and indicate how I disagree with the topic.

Reading skill, undoubtedly, is a very crucial factor to master for every student. As a student who finishes amount of tasks and homework, I must use lots of knowledge and documents that I do not know to make it better. In this case students will read different kinds of books essays, files and so on. Thus, how to read them fast and correctly is a fundamental ability to finish tasks. Though reading skill has many advantages. Is it the most important? I do not think so.

Not only reading skill can help student doing better but also other academic skills will help more. Such as writing, listening, speaking and so on. Firstly, writing skill is a way to show students' views. As everyone know, a paper or an essay must be finished in the end of a semester. So, what score can a student obtain is determined by how to organise words to indicate points. So, in my opinion, it is the most direct approach to gain great academic performance. Then, the listening skill is a prior skill for a students. Before you use the new knowledge, you need hear the content of what professor says clearly. No tutor teaches students only by writing on blackboard without any voice. So, listening is a very effective way to obtain new things. As for speaking, it is a imperative skill for every foreign student to own. Speaking skill is communicative skill. Who wanna be a silent person? Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills. Speaking to others and get more information from others is a correct way to study academy and improve one's ability.

Above all, I think, the most important is not only reading skill. Other skills are not either. The integrated skill is the best way to obtain the biggest progress and a better academic performance.

Appendix 6.7 - Lexical error analysis - instructions for participants

Summary: There are two tasks:

- 1) Identify all lexical and grammatical errors in a piece of student writing.
- 2) Categorise a list of lexical errors using a new framework.

Read the participant information sheet (Appendix A), and if you agree to participate, sign the consent form (Appendix B) provided (pls do this on the data collection day).

Task 1) Identification of Errors

- a) Read the guidance on identification of lexical and grammatical errors below.

Identification Advice

An error is defined as 'a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's highly-skilled user of English counterparts'. (Lennon 1991 p182), 'unless the error is a norm of that variety of English and made in that country' (George 1972, px).

- 4) To identify lexical and grammatical errors, the rules below, adapted from Hemchua and Schmitt (2006), should be followed:
 1. In fixed phrases, such as '*What's *a matter?*' the error is collocational and therefore lexical. Other non-collocation errors with articles (omission, overinclusion and mis-selection) are grammatical.
 2. Derivational affix errors (e.g. *He is kind and considerable) are lexical.
 3. Inflectional affixes (e.g. work/worked/working or dog/dogs) are grammatical. As are genitive errors (errors with pronouns, or adjectives that modify another noun or errors used to show possession). Comparative and superlative affixes are grammatical.
 4. 'Clause errors' are grammatical (e.g. *It's not difficult *for getting to a hospital* and **While waiting, my hamburger went cold*).
 5. Errors in relative clauses (incorrect pronoun and referential error) are grammatical.
 6. Errors in countability, tense, grammatical redundancy (*My mother, she is ill.*) and subject/verb agreement are grammatical.

Separating Lexical Errors

- 5) Multiple lexical errors in a phrase are counted separately. However, when a sentence lacks coherence, and it is difficult to separate errors, count them as one per phrase. For example, in the sentence '**all people there aren't in the globe next a few years*', there are three, based on the three phrases (noun, verb and prepositional phrase).
- b) Underline all the lexical errors and **Embolden** all the grammatical errors that you agree on. Click on record when you are ready to start. Voice your thoughts as you complete the task, speaking clearly. I'm more interested in hearing about issues for debate, rather than areas that you agree upon. Stop the recording when you have finished.

Task 2) Allocation of error to category

- Refer to the framework and guidance (Appendix 3) to understand how to categorise lexical errors. Hopefully you will have had time to read these when they were emailed to you.
- Click on record when you are ready to start. Voice your thoughts as you complete the task, speaking clearly. I'm more interested in hearing about issues for debate, rather than areas that you agree upon. Stop the recording when you have finished.
- Look at the errors below and decide which type of lexical errors they are according to the codes in the corresponding framework cells. Record the codes in the spaces provided
- Allocate a confidence score (one per participant, per error) in the spaces provided:
 - If you are sure that you have correctly and unambiguously coded an error, record a score of '3'.
 - If you have a little doubt over your code, and it could, feasibly have been coded differently, allocate a confidence score of '2'.
 - If you feel the error could have been easily coded differently, allocate a confidence score of '1'.

Table 1 Various Types of Lexical Errors (in italics) and Participants' Confidence Scores

No	Error	P1 Code	P1 Conf	P2 Code	P2 Conf
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.				
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.				
3	perfect academic performan is a <i>basic quality</i> that every student wants.				
4	Let <i>me</i> discuss the topic				
5	and indicate how I disagree with the <i>topic</i> .				
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master				
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master				
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master				

No	Error	P1 Code	P1 Conf	P2 Code	P2 Conf
9	As a student who <i>finishes</i> amount of tasks and homework,				
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .				
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .				
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .				
13	different kinds of books <i>essaies</i> , files and so on.				
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .				
15	an essay must be finished <i>in</i> the end of a semester.				
16	<i>Is it the most important ?</i> I do not think so.				
17	Is it the most important ? <i>I do not think so</i> .				
18	So, what score can a student obtain is determined by <i>how to organise</i> words to indicate points				
19	So, <i>in my opinion</i>				
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .				
21	<i>to gain great academic performance</i>				
22	, the listening skill is <i>a prior skill</i> for a students.				
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.				

24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.				
----	---	--	--	--	--

No	Error	P1 Code	P1 Conf	P2 Code	P2 Conf
25	content of what proffesser says clearly.				
26	content of what proffesser says clearly.				
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .				
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .				
29	Who <i>wannna</i> be a silent person ?				
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills				
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills				
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills				
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>				
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.				
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .				

36	I think, the most important is not only reading skill. <i>Other skills are not. either .</i>				
37	the best way to <i>obtain the biggest</i> progress				

Confidence scores: 3=Very confident about accuracy of allocation. 2= Quite confident. 1=Not Confident.

Table 2 Participants' error categorisations and confidence scores

No	Error	R code	R Conf	P1 Code	P1 Conf	P2 Code	P2 Conf	P3 Code	P3 Conf	P4 Code	P4 Conf	P5 Code	P5 Conf	P6 Code	P6 Conf	% agr	Mode	ave conf
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	41	3	41	2	41	2	18	2	19	2	19	3	19	3	42	41	2.4
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	9	3	9	3	9	3	9	3	9	3	9	3	9	3	100	9	3
3	perfect academic performan is a <i>basic quality</i> that every student wants.	19	3	18	2	18	2	38	3	38	3	25	2	25	2	29	18	2.4
4	Let <i>me</i> discuss the topic	41	3	41	3	41	3	41	2	41	2	42	2	42	2	71	41	2.4
5	and indicate how I disagree with the <i>topic</i> .	19	3	39	3	39	3	18	3	18	2	19	1	19	2	42	19	2.4
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	39	3	1	3	1	2	39	2	39	2	24	1	24	1	42	39	2
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	9	3	9	3	9	3	9	3	9	3	9	3	9	3	100	9	3
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	19	3	18	2	18	2	18	3	18	3	19	3	19	3	57	18	2.7
9	As a student who <i>finishes</i> amount of tasks and homework,	19	3	19	3	19	3	19	3	19	3	19	3	19	3	100	19	3
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	41	3	41	3	41	3	41	3	41	3	41	3	41	3	100	41	3

No	Error	R code	R Conf	P1 Code	P1 Conf	P2 Code	P2 Conf	P3 Code	P3 Conf	P4 Code	P4 Conf	P5 Code	P5 Conf	P6 Code	P6 Conf	% agr	Mode	ave conf
11	I must use lots of knowledge and documents that I do not know to make it better .	27	3	27	2	27	2	17	2	16	2	27	3	27	3	71	27	2.4
12	I must use lots of knowledge and documents that I do not know to make it better .	41	3	41	2	41	2	16	2	16	2	19	3	19	3	42	41	2.4
13	different kinds of books <i>essaies</i> , files and so on.	11	3	11	2	11	2	11	3	11	3	13	3	13	3	71	11	2.7
14	different kinds of books <i>essaies</i> , files and so on.	41	3	41	3	41	3	41	3	41	3	41	3	41	3	100	41	3
15	an essay must be finished <i>in</i> the end of a semester.	22	3	22	2	22	2	22	3	22	3	22	3	22	3	100	22	2.7
16	<i>Is it the most important ?</i> I do not think so.	41	3	41	3	41	3	40	3	40	3	41	3	41	3	71	41	3
17	Is it the most important ? <i>I do not think so.</i>	41	3	38	2	41	2	45	3	45	3	41	3	41	3	57	41	2.7
18	So, what score can a student obtain is determined by <i>how to organise</i> words to indicate points	19	3	19	2	19	2	18	2	18	2	27	3	27	3	42	19	2.4
19	So, <i>in my opinion</i>	41	3	41	3	41	2	41	3	41	3	41	3	41	3	100	41	2.8
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	27	3	25	2	25	2	16	3	16	3	19	3	19	1	29	27	2.4

No	Error	R code	R Conf	P1 Code	P1 Conf	P2 Code	P2 Conf	P3 Code	P3 Conf	P4 Code	P4 Conf	P5 Code	P5 Conf	P6 Code	P6 Conf	% agr	Mode	ave conf
21	to <i>gain great academic performance</i>	27	3	38	2	38	2	19	3	19	3	27	3	27	3	42	27	2.7
22	, the listening skill is <i>a prior skill</i> for a students.	27	3	40	2	40	2	19	3	19	3	18	3	18	3	29	27	2.7
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	41	3	41	3	41	3	41	3	41	3	41	3	41	3	100	41	3
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	19	3	18	2	19	2	20	3	20	3	19	2	19	2	57	19	2.4
25	content of what proffesser says clearly.	10	3	10	3	10	3	10	2	10	2	10	3	10	3	100	10	2.7
26	content of what proffesser says clearly.	11	3	11	3	11	3	11	2	11	2	11	3	11	3	100	11	2.7
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	19	3	19	2	19	2	19	3	19	3	27	3	27	3	71	19	2.7
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	19	3	19	2	19	2	18	3	18	3	19	3	19	3	71	19	2.7
29	Who <i>wannna</i> be a silent person ?	41	3	41	3	41	3	41	3	41	3	41	3	41	3	100	41	3
30	<i>Whatever</i> a student hear and write and how fast a student read, if...	19	3	25	2	25	2	18	3	18	3	41	2	41	2	29	19	2.4

No	Error	R code	R Conf	P1 Code	P1 Conf	P2 Code	P2 Conf	P3 Code	P3 Conf	P4 Code	P4 Conf	P5 Code	P5 Conf	P6 Code	P6 Conf	% Agr	Mode	ave conf
31	... if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	14	3	14	3	14	3	14	3	14	3	14	3	14	3	100	14	3
32	... if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	19	3	19	3	19	3	19	3	19	3	19	3	19	3	100	19	3
33	...if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	27	3	38	2	38	2	17	2	16	2	27	3	27	3	42	27	2.4
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	27	3	3	2	3	2	16	3	16	3	1	3	1	3	29	3	2.7
35	I think, <i>the most important is not only reading skill.</i> Other skills are not. either .	27	3	27	2	27	2	16	3	16	3	40	3	40	2	42	27	2.5
36	I think, the most important is not only reading skill. <i>Other skills are not. either .</i>	27	3	17	2	16	2	16	3	16	3	40	3	40		42	16	2.7
37	the best way to <i>obtain the biggest</i> progress	27	3	35	2	35	2	19	3	19	3	27	2	27	2	42	27	2.4
Averages			3		2. 4		2. 3		2. 7		2. 7		2. 7		2. 7	67		2.6

Key R=Researcher's own analysis

Table 2 Categorisation and Confidence Score results

Appendix 3 Framework for Lexical Error Analysis, Error Codes and Explanation V1

	OMISSION	OVERIN- CLUSION	MISSELEC- TION	MISORDER	BLEND
Section A Formal Lexical Errors					
A1 Suffix Type	1	2	3	N/A	4
A2 Prefix Type	5	6	7	N/A	8
A3 Spelling	9	10	11	12	13
A4 Two words should be one	N/A	N/A	N/A	N/A	14
A5 One word should be two	N/A	N/A	N/A	N/A	15
Section B Semantic Lexical Errors					
B1 Coherence (Words do not exist in English or the whole phrase is unclear)	N/A	N/A	16	N/A	17
B2 Wrong Word (Field Error)	N/A	N/A	18	N/A	N/A
B3 Wrong Word (Sense Relations Error)	N/A	N/A	19	N/A	N/A
B4 Prepositions	20	21	22	N/A	N/A
B5 Cohesion (Lexical substitution, pronoun use, conjunctions)	23	24	25	N/A	26
B6 Paraphrase	27				
B7 Wrong Words (Statistically weighted preferences)	28	29	30	31	32
B8 Arbitrary combinations and irreversible binomials	33	34	35	36	37
B9 Connotative meaning	N/A	N/A	38	N/A	N/A

B10 verbosity/extra word/Repetition	39				
B11 underspecification/ missing word	40				
B12 Formality	N/A	N/A	41	N/A	42
B13 Miscellaneous	43	44	45	46	47

The Categories (Explanation)

Omission	=something has been left out
Over-inclusion	= something should not be there
Mis-selection	=the wrong item has been chosen
Mis-order	=items are in the wrong order
Blend	=two items have been incorrectly combined in some way

Section A

A1 Suffix Type Correct root, but incorrect suffix (e.g. *They were very consider*able* <considerate>).

A2 Prefix Type Correct root, but incorrect prefix (e.g. *It's quite *inimportant.* <unimportant>).

A3 Spelling Meaning is clear. (E.g. *I need to *safe some money.* <save>). Words with unclear meaning should be categorised under B1.

A4 Two words should be one (e.g. **Every one will pay* <everyone>). Moot points should be omitted from the analysis.

A5 One word should be two. (e.g. *Where is my * tooth brush?* <toothbrush>).

Section B

B1 Coherence Words do not exist in English, or meaning of whole phrase is unclear. (e.g. **Where is those that are?<??>*)

B2 Wrong Word (Field) A 'totally wrong word' or Semantic Field Error category. (E.g. *I feel *serious* <stressed>. Compare with B3.

B3 Wrong Word (Sense Relations) Problems discriminating between co-hyponyms, antonymss, synonyms, superonyms, hyponyms and concepts of merynomy: i.e. words in the same lexical set or meaning area as the intended word. Compare with B2 (e.g. *We have modern *equipment in our house.* <appliances>).

B4 Preposition

B5 Cohesion (Lexical Substitution, Pronoun Use, Conjunctions) Also known as Discourse errors. Words or phrases, i.e. as lexical substitution errors when referring to the same subject (e.g. Neymar, he, the PSG striker), incorrect pronoun choice (e.g. He for she) and wrong choice of conjunction (e.g. 'however' instead of 'and').

B6 Paraphrase Could simply be non-expert like phrases or when there is more than one error in a phrase or for whole problematic phrases which require re-writing. Meaning is clear. E.g. **the number of people is really great.* <There is a large number of people>.

B7 Statistically weighted preference Collocational errors relating to numbers or amounts (e.g. *The Army suffered *big losses* <heavy>).

B8 Arbitrary combinations and irreversible binomials Collocational errors relating to fixed phrases with pairs of words (e.g. **hike-hitch* <hitch-hike> **fro and to*).

B9 Connotative Meaning The meaning is clear, but also odd. (e.g. *He's quite *notorious for the charity work he's done* <famous>)

B10 Verbosity/ Unnecessary Words/ Repetition When too many words are used or repeated unnecessarily. E.g. *He bought an apple and* he bought a banana*. <He bought an apple and a banana>.

B11 Stylistic Underspecification /Missing Words Phrases/sentences that require more detail to clarify meaning or improve expression. (E.g. *Although *cars in the country are lower...* <Although there are lower car numbers in the country....>).

B12 Formality Words or phrases that are correct, but seem overly formal or informal for the genre. (E.g. **Kids are not allowed in the lounge after 7pm*. <children>). (E.g. *I *informed my girlfriend via the medium of the telephone*. <I told my girlfriend on the phone>)

B13 Miscellaneous For any errors that do not fit into the categories above.

Examples of errors taken from Zimmermann (1986), Hemchua and Schmitt (2006) and Picot (2017).

Appendix 6.8 Interview schedule for participants using the new framework

1. Number of years teaching experience ____
2. What training in linguistics have you had? Reading or courses?
3. How clear to you was the guidance to use the framework?
Completely clear Quite clear Not sure A little confusing Very confusing
4. If you have answered Quite clear/Not sure/A little confusing/Very confusing, please explain your answer
5. Are the categories in the framework, i.e. the types of error to be allocated to them, sufficiently clear for you?
Completely clear Quite clear Not sure A little confusing Very confusing
6. If you have answered Quite clear/Not sure/A little confusing/Very confusing, please explain your answer to the question above
7. Were the differences between the categories clear enough for you to use the framework easily?
Completely clear Quite clear Not sure A little confusing Very confusing
8. If you have answered Quite clear/Not sure/A little confusing/Very confusing, please explain your answer to the question above
9. Overall, how easy or difficult was it for you to use the framework?
Very easy quite easy Not sure A little difficult Very difficult
10. Please explain your answer to the question above
11. How satisfied are you with the depth of analysis (How detailed are these results?)
12. How might these results be used? (have their results on-hand)
13. Do you have any comments on the guidance or framework?
14. Can you make any suggestions for improving the guidance or framework?

Appendix 6.9.1 P1P2 LEA Transcript

P2 So, *To get a perfect academic performance*. So, I'd say that was about formality (___)

P1 (___) erm (___) get

P2 so formality (___) so, (___)

P1 is it the wrong word?

P2 *To get a perfect academic performance*, but what would the right word?

P1 (___) achieve (___)

P2 OK, but a synonym of achieve would be get, wouldn't it (...) (___) erm, (___) so (___) I'm B2 (___) I'm going to say 41.(...)

P1 Which one is it? Most confident? Quite confident.

P2 Quite confident. I'm going to go with 2 (___). How about you? (...)

P1 yeah go on

P2 How confident?

P1 Errr (___) confident 2

P2 *To get a perfect academic performan*. Ok so that's spelling and omission

P1 Spelling (...)

P2 Spelling is 9

P1 Hmm mm

P2 I'm fairly confident with that so it's 3. You think so?

P1 Yeah

P2 *perfect academic performan is a basic quality that every student wants* (___)

P1 erm (___) basic quality (___) it could be like a wrong word (___)

P2 (___) In the sense relations or the other relation

P1 (___) well, I don't understand B3, so B2

P2 (...) (___) OK So B2 is 18 (___) yeah? (___). And we're pretty sure about that?

P1 (___) yeah

P2 *Let me discuss the topic*

P1 that's the topic. Could be formality?

P2 Hmm mmm So 41 again. And we're pretty sure about that?

P1 yeah

P2 So 3

P2 *indicate how I disagree with the topic*. Is that repetition? Bit that we talked about again, so is that verbosity? Extra word repetition?

P1 yeah

P2 that's 39. Oops pardon. We sure?

P1 yeah

P1 Reading skills (___) suffix? (___) (...) is it suffix? 's' is a suffix?

P2 Yeah? Reading skills. So would you say it was omission? A number one?

P1 (___) yeah

P2 yeah. Not certainly convinced about that. Shall we say 2? (___) are you pretty sure?

P1 I would say 3. I don't know what else it would be.

P2 Fair enough

P2 Undoubtedly, undoubtedly what was spelling number again?

P1 erm A3 it's like something's missed out undoubtedly

P2 Yeah it's omission, so it's 9. Yeah quite sure about that

P1 yeah that's cool.

P2 it isn't a factor. reading skill isn't a factor it's like a skill, so it's the wrong word (___).

P1 (___) which is a (___) B2 or like B3. Could factor be a synonym? A wrong synonym used maybe?

P2 factor would be better. Reading skills are undoubtedly a very crucial (___) Reading is undoubtedly a very crucial skill to master. If you didn't have 'skill' before. Skill isn't a factor, is it? (___). I'd say it was probably a field error (___).

P1 so B2

P2 yea? so 18. Are we sure?

P1 Err. Probably say 2

P2 *As a student who finishes amount of tasks and homework*. So this is B3. It's a synonym, yeah?

P1 Yeah.

P2 And so that could be 19. D'y' reckon?

P1 misselection yeah.

P1 *I must use lots of knowledge* so that's formality

P2 yeah

P2 *I must use lots of knowledge and documents that I do not know to make it better*. Do you think he means like articles? Research reading. documents

P1 erm.. *I must use lots of knowledge and documents that I do not know* (___)
documents that I do not know (___)

P2 It's all the same

P1 *documents that I do not know* (...)

P2 can we use *do not know*. when I read that I thought he's kind of saying that he had to read new things that he hadn't read before like if he had used 'new articles' that he doesn't know now. That's what meant new things

P1 yeah

P2 *I do not know*. (___)

P1 Maybe paraphrase?

P2 Yeah I'd say so cos I don't know what else it could (___) yeah? How confident?
Hmm 2?

P1 Yeah probably 2

P2 (...) *to make it better* so 'improve' would be a better way of saying it

P1 erm (___) would you say that's formality? Or?

P2 Yeah probably

P2 *different kinds of books essaies* spelling

P1 (indistinct) one of the easy ones spelling

P2 spelling because its

P1 essays he's included something. Overinclusion

P2 but it's almost like he's used the e. um changing the y to ies. It's like he's overused a rule or used a rule inappropriately because if it was like babies, It'd be correct (___) so should that

P1 like something should not be there or a wrong item has been chosen.
Misselection

P2 Misselection I think. It might be that. Yeah?

P1 11 yeah?

P2 (...) I know it's definitely spelling but um

P2 *different kinds of books essaies, files and so on*. Formality again?

P1 formality (___) misselection? (...)

P2 *an essay must be finished in the end of a semester*. This is what we were talking about earlier. In the end of the semester. At the end of the semester. So it's preposition problem but because of the phraseso I think it's preposition misselection. I think it's 22.....what do you think?

P1 (___) I'd say (___) where's er B4 (___) er yeah (___)

P2 *Is it the most important ?* So it's all about *I do not think so*. I'd say formality

P1 Yeah because it's rhetorical

P2 are we sure? I'm sure

P1 (___) yeah it's 41

P2 yeah 2 or 3?

P1 I'd put 3

P2 And what about this one? *I do not think so.*

P1 (___) is not like words is that redundant (___) hmm (___) too many words

P2 the only thing is B10, but it's not that. It's not repetitive or wordy.

P1 mmm you could say it's unnecessary words but

P2 it's like an unnecessary phrase really, isn't it?

P1 Or B9 in terms of like it's odd that he would put it. Coz you wouldn't normally answer.

P2 or B8 or B9

P1 Connotative meaning B9 It's clear but also odd. It's odd to answer a rhetorical question.

P2 That's true. That's a good point. So that would be 38.

P1 er B9 yeah..I don't see where that's (___) (...)

P2 What do you reckon. Shall I put one and you put the other? (...)

P2 Er yeah

P2 So, what score can a student obtain is determined by how to organise words

P1 Is that how to organise? Is that like choice of words? I don't know how to categorise. So, what score can a student obtain is determined by how to organise words (...)

P2 so maybe that's the (indistinct) one B3. I'd say it's be that one because it's

P1 errr (___) I think yeah

P2 What's that again? B3 19

P2 So, in my opinion. That's formality. 41?

P1 yeah

P2 It is a most Direct approach (___) It is a most Direct approach

P1 (...) looks like this (___)

P2 I'd have said that was grammatical in the end. Doesn't make sense

P1 Could it be like cohesion?

P2 Ah yeah good point. Lexical substitution so misselection

P1 B5

P2 25 Yeah a wrong item has been chosen. If we decide really quickly it's a 3 (...)

P2 *to gain great academic performance* So this was a problem with gaining, then wasn't it (___)

P1 yeah (...) erm 9

P2 Maybe it's a connotative meaning. B9. I can completely understand what the person is saying: means to do well in academia. But it's not quite right how they said it.

P1 mmmm Or it could be like B6 maybe. But that's might how you might say it.

P2 mmmm

P1 Wrong item has been chosen.

P2 But if we were trying to decide between two, as Tony said, we should go with the one that's more specific which would be the connotative meaning 38 (___) Are you going to go for that as well?

P1 yeah

P2 *A prior skill for...* So they mean a skill for...a prior skill. What do they say that they want to say?

P1 Prior. Weird how they used prior (___) Could be B2 or could be B1 except it's not (___) is a prior skill. Ah. I suppose if you use prior

P2 I don't know if it's a B3 because it's like in a

P1 B11 sentence requires more detail to clarify meaning or improve expression

P2 yeah that would make sense as we don't really know what he wants to say. So B11 was number 40?

P2 *Before you use the new knowledge, you need to*

P1 Formality?

P1 Hmmmmm

P2 *you need hear the content of what professor says* so I guess that they're using synonyms so I think that that would be B3 so wrong word from (...) because it should be listen to.

P1 listen to or would you say understand. You need to understand the content

P2 But either way it's still related in a sense. It's not completely the wrong word in terms of it's a different like idea although it could be I guess. Understand and listen is different from listen and hear

P1 Hmmm

P2 (___) but it's not a totally wrong word is it? because we understand what he means.

P1 Different in terms of field if you listen to something. It's totally different if you understand it

P2 Hmm Shall we go for a different one each.

P1 yeah go on I'll say B2 18 that's a 2

P2 OK. *professer* Double so it's spelling but what kind of spelling?

P1 Well you've got the (...)

P2 But there's two of them. That one's the O There's two different spelling issues there's that spelling issue, the F. and there's the errr (...) there's overinclusion, number 10 and then wrong letter, misselection is 11 for the next one.

P1 Yeah

P2 Ok it works. Pretty sure about that one I think.

P2 *No tutor teaches students only by writing on blackboard without any voice* (___) without a voice. (...) Without speaking is probably what they need (___).

P1 erm (___) So, you could say B9? Could say B6

P2 OK. It's B9 I think. A bit odd.

P1 But is it connotations, though? Is it connotations? Without a voice, without speaking (___) wrong semantic field error (___) so I feel like it's the same: like voice and speaking, but it's. I don't know if it's a different term.

P2 (indistinct) could be B3

P1 yeah same meaning. I'd go for B3

P2 So 19?

P1 yeah

P2 I'd go 2

P2 *As for speaking, it is a imperative skill for every foreign student to own.* So that's the wrong word. Just wrong. haha Field error? Own or have. We want it to be have. You don't own a word so that would also be B3.

P1 Would it though

P2 Well own is a synonym of have. Have means so many things. But in that, you can't own a skill. You have a skill (___).

P1 That's not a fixed phrase then is it? like in B8 where you've got like collocation error.

P2 No I don't think it's a (...) error. I'd say it's 19 again.

P1 Well what's that statistically weighted preference?

P2 More about the amount of something

(...)

P2 Would you agree?

P1 Yeah go on.

P2 *Who wannna be a silent person ?* That's formality again.

P2 This is a hard bit *Whatever a student*

P1 (...) *Whatever* (___)

P2 (___) What did we say it was meant to be?

P1 Regardless of (___) so is it like coherence?

P2 yeah maybe misselection coherence B1

P1 Meaning as a phrase unclear whatever. Regardless of

P2 I don't think it's completely unclear though.

P1 I think B3? Whatever, regardless Or wrong place of conjunction, cohesion. B5.

P2 So Ok yeah so that's misselection 25?

P1 Yeah wrong item B5

P2 You could just slip regardless in there and it's fine for a bit.

P2 OK *Whatever a student hear and write and how fast a student read, if a student can not so can not* is the problem there, so it's two words should be one

P1 yeah

P2 14?

P1 14 yeah

P2 Not talk, say. Should be say their own views, shouldn't it? So that would be B3
19

P1 B3 19 Misselection. Yeah

P1 *it is vein for the student* so could that be like (___)

P2 I don't know what he means there (___)

P1 What are they trying to say if he can't discuss his own views (___) it doesn't matter how many good skills you've got.

P2 OK

P1 So it's kind of like the meaning is clear but odd. Vein. Does that even exist in English?

P2 OK. Do they mean that it is in vein. Like it doesn't really matter. Is that what they're trying to have? Maybe they've found that fixed phrase but they've used it badly. So having so many skills is in vein. We wouldn't use it like

P1 It would be odd so you could say B9 In vein is used in bit and that's also spelt wrong

P2 So shall we say B9 because we could just say paraphrase as well coz the whole thing is a bit wrong

P1 Well yeah but it's not like a non-native phrase. I feel like it's a native phrase but they've just misused it

P2 So B9 then. 38

315

P1 A B9 yeah

P2 Speaking to others and get more information from others is a correct way to study academy study academy I don't know what that is. I don't know if it's academically

P1 Speaking to others and get more information from others is a correct way to study academically. In which case, is that suffix?

P2 And is it Missselection 3?

P1 yeah?

P2 Yeah?

P1 but probably 2

P2 (...)

P2 *I think, the most important is not only reading skill.* Confusing sentence

P1 *I think, the most important is not only reading skill.* (...).

P2 Coherence? The whole phrase is unclear cos it's that whole thing.

P1 Erm (___) (...) (___).

P2 Coherence or paraphrase?

P1 Well but the meaning is clear so I would say the meaning is clear in that but it's the rest of it. The most important is not only (___)

P2 But if you have that as a sentence on its own, the meaning isn't clear (...)

P1 But wouldn't you assume that it's going to carry on (___)

P2 Yeah but the fact that it doesn't means that the meaning isn't clear. Cos that means it meant something else.

P1 Yeah but then when you get to like this bit.

P2 Others are not either

P1 Like that would be the way it carries on.

P2 D'you think it's a paraphrase error then. You'd just need to re-write it.

P1 I don't know. Like comparing those two, I'd say the meaning clear in that bit but other skills are not. This bit is not clear. Because it's not

P2 Ok so if it is clear do you think that he's just added this word that he shouldn't have? (indistinct)

P1 No because you would say that, wouldn't you? It's not only this but

P2 Yeah, but he doesn't go on to say but.

P1 Yeah but that second part, the meaning is unclear

P2 Do you remember in the essay, that wasn't what he was trying to say. He's saying reading skills aren't the most important and others aren't the most important. He's basically saying there's no most important.

P1 Yeah but like the first bit was OK. It's the second bit that was

P2 OK so what's wrong with it?

P1 Well, I'd probably say the meaning is clear but you'd want to rewrite it

P2 So paraphrase

P1 Yeah, so paraphrase

P2 I think so cos they said

P1 27

(...)

P2 I'd say that the meaning is not clear and you'd need to rewrite it.

P1 Yeah so there would be coherence

P2 OK

P1 B1

P2 *the best way to obtain the biggest progress*

P1 Would you say 16? What's 17 two items have been incorrectly combined.

P2 We don't know what the two items are

P1 Well there. Other skills are not, either

P2 and P1 (laughter as they change the previous P1 code to 17)

P2 the best way to obtain the biggest progress. What's the problem there? It was this word, wasn't it?

P1 The best way to make the best progress?

P2 So what shall we say? Wrong word? Sense relation? (___) So basically there's two wrong words. One is completely wrong. The other is you're in that area. Like maybe the choice of word isn't quite appropriate.

P1 But's what's that like collocation? the second word would collocate with progress.

P2 But then it says relating to fixed pairs. Pairs of words. But I suppose you could say make progress, couldn't you. That's a collocation. You don't obtain progress
Yeah maybe it is that one then. What was that number?

P1 Err B8

P2 Lots of options there. Erm So is it that the

P1 Would you say 35? Misselection?

P2 So it should be make progress It's also that. You can't say big progress (___)
significant progress So It's a collocation issue.

P1 mmm

Appendix 6.9.2 P3P4 LEA transcript

P3: This is academic writing, isn't it? We think so. Reasonably perform. Yeah. Let's get a perfect academic performance. We have to ignore that because it's spelling

error mostly.

P4: Is it is it? No. Is it about the it's,

P3: oh, you mean they've got the wrong derivational affix as though they heard performan instead of performance or is it that they think that is the derivational affix for perform?

P4: Yeah. They think that'd be a unique one, wouldn't it? Yes. But maybe because there are other potential affixes aren't there. But they picked. Yeah.

P3: Okay. Well, let's be super strict. Let's call it lexical then. Although sense it might, Yeah. Yeah. They just hadn't heard.

P4: They haven't heard the ending, have they? It's like, yeah, I mean, it's like, it's like you would say if it was an ED that wasn't there, except that would be grammar.

P3: Yeah, It's definitely closer to derivational than affix then Yes. It's not a sorry, where's the underlining area? Where is it? Actually,

P4: it's not there, but I think you can do control and U. Okay.

P3: Let's see what's there is a basic quality that every student wants. Quality is a lexical mis choice, I suppose.

P4: Yeah, It's not a quality. It's not a quality. Yeah. You can see it. Yeah. Again wants. It's not right but it's kind of right, isn't it? Yeah, it's not wrong,

P3: It's not very formal. And you think I was thinking, get is not really the most skillful you would attain, achieve, or something of the..

P4: and you don't get a performance, do you not perform?

P3: If you're thinking about acting, you have to achieve it. Perhaps performance might also be the wrong one. I think we're right to underline that. We also underline get, I think it's not the ideal.

P4: No, because you just wouldn't say that, would you?

P3: Students must develop various skills to achieve the goal.

P4: A variety of. No, Maybe it's okay.

P3: Perhaps that's not too bad. Not especially formal, but is as clear as if you were chatting. Yeah. You'd say that there are many skills to help the students to study. It's not skills more like strategies or tactics.

P3: Yeah, Yeah, I think that's just the wrong word.... Isn't some people think the reading skill, the reading skill is the most important compared with others?

P4: Well, yeah, just about. The skill of reading would be better. Reading. Just reading without is the most important skill. Yeah. You'd really have to recast it to make. Should we leave it? Because the is not the problem

P3: we're supposed to, that's a grammatical problem. Then you'd have to get rid of skill as you say. It's the most important compared with others. Stylistically but it's clear. However, others hold the opposite view.

P4: It's not opposite, is it? Because there are four that yeah,

P3: There has to be a different view, but is that a lexical error or semantic one? Is it that natives can make the same mistake if...

P3: yes yes, it's not the wrong word. It's the wrong idea. It's okay.

P4: Let me discuss the topic. Let me discuss the topic and indicate how I disagree with the topic. Well again, conceptually, it's not the topic he's disagreeing with. It's the viewpoint. Yeah, it's not about the topic. Topic, it's the wrong word.

P3: It's the topic is the wrong word, isn't it? I think I think there's an overlap there with concept. Yes, But we'll underline it. Having said that..

P4: It was the same, isn't it? It's the same problem twice.

P3: Let me discuss the topic. That sounds better to me than disagreeing with the topic because the topic can be a more general subject.

P4: Yeah a more general discussion.

P3: Okay, fine. To me, that's more (...), Now we're doing reading skill without the reading skill and it's still not right. As you said before, reading, undoubtedly the spelling... is a very crucial factor to master for every student. What we gonna do with this redundant skill?... we've introduced that word skill in our teaching to them. Otherwise a normal person would've put 'reading'. They've never learned academic writing.

P4: They wouldn't consider it as a skill in the same way,

P3: That's an introduced problem for them funnily enough. But I don't know what we do with redundancy. That's a good question because...

P4: yeah, if we're doing it, we have to do it back there as well. It's the same problem, isn't it? A lexical form or a combination of forms which would not be made by a highly skilled English user.

P3: Oh, okay.

P4: See, I think that helps us, doesn't it? But I think people might say that highly skilled, you wouldn't say reading skill, but would you? Reading is the most important skill. Yeah. So you would use skill but not like that in a collocation. It's wrong because...

P3: you're saying it twice. It's a redundancy thing Yes. Isn't it? But unless you do another thing in the skill of reading,

P4: but also Yeah. I think generally you talk about my reading skill. You say my reading is good or my reading is bad. or

P3 Shall we underline it?

P4: Yeah. I think so, yeah. it's it's not being used properly, so Yeah, that's not a correct combination.

P3: I feel it might be in many languages, but it's not done in English

P4: and it makes sense, but it's not....

P3: So you said we've called it your skill. Yeah.

P4: We've probably contributed there.

P3: is a very crucial factor to master. Is a reading skill a factor?

P4: I don't think is a factor? Wrong word picking. Are we? Yeah, that's right. But are we picking on very crucial. Ah no. that's a..

P3: countability.

P4: Yeah. So we're not, we as a student who finishes amount of tasks in the homework,

P3: finishes?

P4: if they complete...It's just style, isn't it?

P3: it's a Germanic word, when we like the Latinate one. Tricky. is an amount of tasks. That's simply a article error.

P4: Yes. But an amount of tasks. A number of tasks?

P3: Well, that's countability then tasks, so we have to leave it. It's like less and fewer.

P4: It's like some tasks. but I think it's..I think amount is the wrong word because we're acknowledging that there's more than one task.

P3: But isn't it the same as less and fewer?

P4: I don't think it is. Because I think that's about much and.. that's grammar, isn't it?

P3: Yeah,

P4: but this is the wrong word because you don't say an amount of tasks. You might say a number of tasks. So the of tasks is okay. The fact that it, it's countable is okay, but we just don't say...

P3: no. Amount is not a non count and number is count, that's the difference. An amount of flower, a number of children, I think if we're letting them off less and much, we have to let them off amount and number for me. I agree. You don't say it. And

P4: and homework.

P3: An amount of homework, of course. Yeah. It does work. I must use lots of knowledge and documents that I do not know to make it better.

P4: That's a comma. Knowledge and documents that I do not know.

P3: Okay. So punctuation we can leave.

P4 to make it better..

P3: I don't quite understand what that means.

P4 I think it means I have to use lots of knowledge and documents that I'm not familiar with to improve the tasks and homework that I do. But that's not what it is. What it says.

P3: Yeah, partly again, these Germanic phrases like I must use lots of knowledge rather than it's important to incorporate a great deal...you know, you can imagine a more formal register would change how it sounded. In this case, students will read different kinds of books, essays, files, and so on. That's okay.

P4: Yeah, I think that's okay.

P3: Thus, how to read them fast and correctly is a fundamental ability to finish tasks. That's a grammar error in order to get

P4: thus

P3: quite close.

P4: Yeah, if this finishes I don't like, I mean, I know we say, but like here and here you want.

P3: You want it complete.

P4: Yeah. But it's only because it sounds funny. But then a highly skilled person...I finished, we finished, I finished the task. Yeah. I mean, if someone says to you, have you finished that task so there's a student doing an exercise, have you finished that task?

P3: Yes, I finished means it's all done. But if you say you need to finish this form in order to get the transport, it's completely wrong. And that's the mistake, isn't it? So let's say it's lexical, really? It's more like completing a form. A mistake, isn't it?

P4: Yeah

P3: let's be strict then. I think that is quite strict, isn't it? That one in the dictionary. In the thesaurus they're given as the same, but we know the usage is not.

P4: And you'd just say style, wouldn't you? I'd say that was style.

P3: It's on that cusp isn't it? It's not a misunderstanding of the word. They understand the word and we understand what they mean, but it does sound wrong.

P4: The reading skill has many advantages...

P3: That's grammatically wrong,

P4: But reading skill doesn't have many advantages, does it?

P3: conceptually wrong also.

P4: I think advantages is not the word, Are we allowed to?

P3: But you can't get it right by just changing a lexical item, can you? Because it's the concept. The whole thing?

P4: Yeah. Yeah. The whole thing is wrong, isn't it? It's the ability to read has many benefits.

P3: Yeah. That's how you'd recast it. the whole thing, But is it the most important? I don't think so. That's So shall we just leave it....because it's not only lexical.... is not only reading skill can help students doing better, but also other academic skills will help more. We've got reading skill, this keeps coming, doesn't it? That's a grammatical..to do better. So that's, yeah,

P4: it's style. There also will help more. But, of the but and the also? but then...

P3: that's a bit more grammar than... then again, you've got a non-sentence, such as writing, listening, speaking and so on. But yeah, we know what they mean. Firstly, writing skill...

P4: Again, same... is a way to show students views.

P3: Yes, as everyone knows, again, grammar, a paper or an essay must be finished. We've done that one already. Do we redo the same.....?

P4: We haven't done reading skill again. So maybe we, because that could work though. It must be finished by the end of the semester.

P3: Meaning, finished in that sense?

P4: yes, I must have finished writing it.

P3: Oh, okay, it might be okay. A preposition error?

P3: I don't think. Oh, hold on. If it's not a phrase..., there was something about...

P3: So it's collocated with the by the end of in the end of...

P3: clause errors? No. Doesn't mention prepositions, actually. So maybe we can have it

P3: So shall I just under line just in because we want by.

P4: Yeah. Because actually it could be at, couldn't it? Yes, it's true either. We don't know.

P3: What score can a student obtain is determined by how to organise words to indicate points? That is not quite clear, but it starts off well, well, actually that's

grammar error, but it's determined by how to organise words to indicate points. Is that to achieve points, do you think? Oh to demonstrate points to make clear.

P4: If your essay is clear? So the thing is that score and points here aren't the same thing. The score is your mark. And the points of the ideas in your essay, I think.

P4: Uh huh. Yeah.

P3: The score a student gets is determined by how they organise the essay. To, to explain,

P3: get their points as

P4: indicate we demonstrate. I don't think it's indicates. I don't think it's...

P3: in in my opinion, it is the most direct approach to gain great academic performance.

P4: Direct, yeah.

P3: Then the listening skill is a prior skill for a student. So is that then that's the wrong word, really, isn't it?

P3: Yeah. Because you'd say in addition, however, or...

P3: it might even be the thus that he used before prior.

P3: He means an a priori skill skill, you know, prerequisite? Yeah. Yeah. Yeah, Again, it's close to the right word. Shall we underline it, then?

P3: I think so. I just don't think, yeah, he knows what he means. But if you read that as a native and you just read it, you'd be like... not right. Before you use the new knowledge, you need to hear, grammar, the content of what the Professor, spelling, says clearly. You need to hear the new knowledge clearly. That's what he's saying. No tutor teaches students only by writing on the blackboard without any voice. Listening is a very effective way to obtain new things. You learn new things. It's like (.....) style. It's not obtained. You know, in some languages that word would be the same, but it's not in English.

P4: is this voice? I know what they mean. I understand but it's not without any voice. I mean, it's or was it just badly phrased? I mean, it's wrong. You wouldn't as a highly skilled user, you would not say that the meaning is clear.

P3: without using his voice? Is that what we would do with

P4: Yeah, kind of without speaking? I suppose he's trying to he's trying to give another word rather than saying. Speaking skilled speaking skill, isn't it?

P3: Yeah. Without using his voice. But you wouldn't say without any voice, even in a normal in any other construction, would you? He'd say he doesn't have a voice or something, but I think it's a structure. It doesn't work.

P4: So do the whole phrase. I think so, I mean, without a voice. Yes. But

P3: listening is a very effect way to obtain new things. As for speaking, it is an imperative skill for every foreign student to again own. In some languages,

P4: yes, yes, yeah. You can see how it's been translated.

P3: Yeah, it's like a transliteration. Slightly. The speaking skill is a communicative skill. Okay. Who wanna be a silent person? That's peculiar.

P4: Is that wrong because that's spoken? That's not...

P3: it's a gross register error, isn't it?

P4: Is it an incorrect lexical form.

P3: I think it's not in the realms of a lexical error. We'd understand him, if he was speaking to us. It's not the right thing here, but we've heard it (...). It's funny spelling too. whatever a student hear and write and how fast a student read. If a student cannot talk own view with others, it is vain for the student to have so many good skills. I know what he means. Whatever a student hears and writes and however however, I think it's not quite the right one. Yeah. How fast?

P4: Okay.

P3: Cannot talk his own views with others say express his own.

P4: Express. Yeah, it's not talk. I think that's he could have talked about. Yeah. It is vain for the student. If he spelled it the other way, it is in vain

P3: but it still doesn't, no. Yeah, I know what you mean, but it's not right. I think we'd underline it. Yeah. In real life didn't have so many good skills. I have so many good skills. Speaking to others and getting more information with others is a correct way to study academy and improve one's ability academically. So it's a derivation error, so that's one we can have.

P4: Way? Yes. Correct. Correct. It's not a correct way.

P3: Just far enough off. I think this is the whole thing, isn't it? Because you don't say a correct, it would be useful way or a positive way, or something. Overall, I think the most important is not only reading skills, other skills are not either go to pieces. The integrated skill is the best way to obtain the biggest progress and a better academic performance.

P4: Can you say that? Yeah, it's kind of stops in time. But the most important...

P3: lexically though, I don't think it's really wrong there. Big progress was not, you know, we would usually say greatest, obviously.

P3: So that's the only thing. I think the biggest progress. Again, it's a register. .. it's a register because if you're chatting, I made big progress.

P4: I wouldn't say that. It's a lot of yeah. it's clear though, isn't it? The meaning, the meanings there?

P3: to get a perfect academic performance is a basic quality, so we're focusing on the get. Yeah. Because he's italicised it. Yeah. Right. Single word.... Oh, no.

P4: Oh, is it B2g? That's not two. No, there is a single word. Yeah, it is. Just the get isn't it? No. Yeah. So the last one of that categories is verb,

P3: what about this first B2 one with a single word could be related or unrelated to correct word. Word exists in word exists in English, does not fit into. So it's that first B2.

P4: Do I think it's a verb? I think I mean it is a verb, but I think

P3: Oh, I see. Is that heading for all of these? Sorry. Yeah, I think sorry. Yes. B2g then, yeah. Does that because it's a verb?

P4: Yeah. Yeah. All right. And it's not it's just one word so..

P3: B2G?

P4: No, we need to decide if it's which. So Missselection.

P3: Oh yeah. MissSelection. If it's chosen the wrong word. Yeah. Sorry. It's I'll put 46?

P4: I think so. Yeah. And then how do you feel? And then how do you feel about that? So this is your so the one next to it, the confidence. So you three and I'm four. I can be okay. I'm saying three, but we both.

P3: Yes. I think I think that's fine. I agree with you. Okay. All right. This was one that was slightly puzzled us.

P4: Oh, yeah,

P3: This is So I think it is important whether or not this is just a spelling he forgot to put CE or whether he thinks the derivation is performan, so is it suffix, right? For me it's a derivation or suffix more than a spelling. But yeah, there are two letters missing, you could say. I really think that was tricky but

P4: Well, I think I would probably say derivation suffix where they are, so that very first, is it A1: Correct root but incorrect suffix and it's not to do with tenses. It is derivation. Yeah, I think I would say that but I'm going to put myself on a two or something.

P3 Does that include...? Yeah, I'm going to put number 3, 2. Confidence. Personally, I think there's any other way of looking at it.

P4: I'm wondering about the spelling but we just don't know, do we? I mean, like I said that they I think I'm going to go for the spelling because I think they maybe haven't heard it. We don't know. They do that. They've heard performance and haven't heard the ending and so that's how they think it's pronounced.

P3: They think that performance or we're not meant to speculate, but that's the singular performance. Is the plural probability maybe. But we're not meant to speculate on.

P4: All right, so you're going for a three. I'll go for a misselection.

P3: Yeah. 03. Am I number three. I've forgotten. I can't remember which way amount we are. I think I'm number three.

P3: So I put me in three because you were thinking something differently.

P4: Well, yeah.

P3: So you think more of the spelling error? Yes, so I missed that.

P4: Is it Omission then that they've missed the end off the word because they don't know how to spell it? Nine....., blend does not.

P3: The words exist in English. It doesn't does it?

P4: Then that's not spelling, is it? Because, But look at the example. Put the folwers in water that I would say the meaning is clear. The real word exists in English. It's not. They're trying to spell a real word. Just made up completely pretend word. I'm going, I'm going to go for nine because I think they've just missed the end of. I'm going to say two.

P3: Yes, that sounds good. Okay. Now on basic quality, a perfect academic performance is a basic quality that every student wants. It's a sort of phrase that

P4: this is more than one word, just not the right phrase. Phrase. Yeah. Yeah. Okay.

P3: I'm pretty confident about that one. Yeah. Basic quality misselection 49. Yeah, I think so. Okay.

P4: We discussed the topic. So that's that's style, isn't it?

P3: I thought that wasn't really wrong. It's not even register, is it? Because the registers All right. It's just we don't quite say that,

P4: but it's not the right register for an essay.

P3: No. So it's all about the writing, I suppose,

P4: because yeah, you could say it. Yeah, you wouldn't quite say it. But it's not, is it not quite formal enough? Is that it? But would you say you'd say, I'm going to discuss the topic or the topic will be discussed? You wouldn't say, let me, would you? But then it's the whole thing that's wrong. It's not the Let me that we're focusing on, do you So I mean, I think it was just to let me that would be a different thing

P3: and I still feel that if the person said it in a speaking context, it would be unnoticed. But it is register. Or is it? I think it is then.

P4: Because if it sounds all right to us in speaking, I think you'd get away with it. Yes. Yeah.

P3: Perhaps not the best, but you know.

P4: Yeah. You almost wouldn't notice it was funny, would you? So is it formality then

P3: some are going to a 55. Or is it more error like, you know, a B3-ish type of thing. It's near 55.

P4: I think if it's the whole phrase, I would go with the 55. I think it was about the Let me then I think I might choose something else.

P3: Yeah. I'm only going to put two because I feel a little bit queasy about formal because that's just the nearest thing I can find for style. Yeah, now we were struggling with this one, I think, and indicate how I disagree with the topic because we were going conceptually, it's not the topic you're disagreeing with. You're disagreeing with that position on it. It's the word choice. Word choice, meaning.

P4: Is there a meaning meaning section B noun B2b? It's just the wrong word.

P3: It's just the wrong word. So you think MissSelection? 27?

P4: Yeah. Yeah, that sounds good. 27. And I think that's exactly it. It's Missselection rather than, I don't know the word or I could have used a better word. They've just chosen the wrong, they don't quite know what it means or...

P3: do you feel number three about that? I'm personally a bit number two ish, but okay, we have different. Okay. The reason I'm just slightly more hesitant is I think a native, I think I've read things like this by native speakers as well, but they're conceptually not spot on. You find allreading skill, I'm pretty sure it's just spelling.

P4: We're on skill, I think.

P3: Oh, sorry. Reading skill, Yes. So that's when we talked about how over, over, over inclusion of three then.. Is it B2? it's just one word, isn't it? It's a single word. that's right. But Noun. Noun. Okay, so and then overinclusion, do we think 26? Yes, it should just be one word. But what he's put is more should say 26?

P4: I think so. Yeah. I'm happy with it.

P3: And then the undoubtedly is purely spelling.

P4: That's at the top, it's a three.

P3: Just a misordering of the letters. Is it? Doubt a little bit missing actually. So it could be nine.

P4: Yeah, nine, yes. Yeah, a factor. All right.

P3: And we noticed that was just not the word, the right word.

P4: It was the collocation, wasn't it? It was the crucial factor even though we understood. Yes. So is it now and again, as we were before, B2b. 27 misselection. Yeah.

P3: Again, this finish is very similar to complete, but not right.

P4: See, that's kind of style, isn't it?

P3: But we haven't got style. I think we'll say the Missselection.

P4: Yeah, it's meaning then, isn't it meaning? It's not B type thing. Yeah, it's not coherence. Is it?

P3: It was probably B2. Is it the error with a single word?

P4: Yeah, that's what I think. Because it's not it's not the wrong word, is it?

P3: No, it's just a slight slightly not right in this context.. So 19, Is it?

P4: Miss selection? Yeah. Yeah.

P3: That's and then there's lots of where, again, bit informal. I think that's more or less a formality issue. We know what it means. Is it a B3d?

P4: Yes. I think it's just a 55.

P3: Okay. Must use those documents and documents that I do not know. Oh, sorry. Lots of knowledge and documents that I do not know.

P4: I don't know which documents. It's not it's not how I understood it. Okay. So it's more than one, it's a phrase problem, isn't it? Is it Lack of clarity, unknown documents. Well, that's the thing, isn't it?

P3: I don't quite know under specification. Isn't it rather useful, more detail to improve the expression?

P4: Because it is it that we don't know about them? Is it that we don't know which ones they are? We can guess the meaning but we don't know exactly what they mean, do we? 54, meaning is unclear. Categorised as B1, coherence. It's this isn't it 50? 54. Much closer to 50? Yeah. I think because Yeah, we've got a sense but actually we don't know which way they what exactly they mean.

P3: I'm going to put two just because of my, you know, But you want a Three or two?

P4: Yeah, three. Feel confident. 33 extra confident!. I must use lots of knowledge and documents that I do, not, that I do not know to make it better, even more obscure actually. Because. Make what better? The writing. I suppose that's the same one. Make better what?

P3: I think I'm quite confident. Yeah, that that's that type of thing make it. Just trying to be careful. Okay. Now, this is purely spelling error.

P4: So that's A3, but ..

P3: meaning is clear, word exists.

P4: Misselection?

P3: Yes, he puts the wrong letter in. Yeah. It's not missorder because he spells it correctly. A bit later on. Actually, 11. 11, okay. So actually, I don't know that I thought that was too bad, but it's formality bit. Informal formality 55. You certainly see native speakers doing that. They must be finished in the end of a semester. So just preposition I suppose.

P4: Yeah, because it's meaning, isn't it? So B2, time and place, and dependent prepositions 38.

P3: Is it the most important? It's another one of those ones. 55 type thing. I would say we're not just quite clear what he means. Oh wait that's formality

P4: 54 under specification

P3: 54 rather. I mean the meaning is unclear in one sense because we don't know to what it refers but it's, it's not mysterious, is it?

P4: It's not nonsense. 54? Yeah. Yeah. I think that's right. I don't think there's anything else that can be. So I think that's for formality 55.

P3: 55, yes. What score can a student obtain? Is determined by how to organise words to indicate points. He tried to do a synonym for how to write. Is it just the wrong choice of noun if you put how to express yourself or something, it would be all right.

P4: But see the words isn't included it because I was going to say it's organised words, but it's not.

P3: It's how we don't tend to use organise with words. So collocational defect, maybe.. We do organise our thoughts though.

P4: Yeah, that's the trouble, isn't it? But it's not how or how you organise. If it was organised, how you organise your thoughts or how thoughts are organised.

P3: Phrase taken as a whole then is just not right.

P4: Yeah. Because we don't say how to organise.

P3: It might be the use of words or something completely different. So if we've got one for the whole phrase is just misselected

P4: Phrase error. Meaning is clear. But word choices ah, B3 word choices are non native like.

P3: Yeah, that's quite close, I think.

P4: Yeah, 49.

P3: Now, again, in my opinion, I don't know if I was particularly harsh. That's the old according to me, Do we feel it's just a little informal?

P4: Yeah. I don't think it's bad. Really? Apart from that

P3: border line, acceptable. Isn't it really

P4: I mean, I wouldn't accept it, but I would understand what they mean and there's no problem. It's not their worst by far.

P3: So, In my opinion, it is the most direct approach. That's more error strain, isn't it? To gain great academic performance, it is the most direct approach. It's the best way. It's not far off actually.

P4: Phrase level error misselection? It's not an approach, it's not that's a direct approach. It doesn't sound right to me.

P3: Also used for whole problematic phrases which require rewriting, so Okay. I reckon 49. Yeah.

P4: Okay. Yeah. I'm not super confident but I'm not sure what else I would say.

P3: 2 then, so that it's not all 3, 3, 3. might give a useful data to gain great academic performance. It's got the grammar omission as well. To gain a great,

P4: it's not right is gain, isn't it? Great academic performance. That's just nonsense,

P3: I think I know what they mean. But, but it's all wrong, isn't it? Although all the words are wrong, we still understand it's not coherent, Is it?

P3: Selection?

P4: Probably phrase, phrase level, it's more than one, isn't it? It's all wrong. The gain is wrong, the great is wrong. I suppose it's phrases. (....) Yeah. Phrase is it 49 again?

P3: Do you meaning is clear, but word choice is yes. Yeah, I think it is. Because it is clear. It's not mystified by what they mean, it's none of the choices quite right.

P4: It's interesting because if you gave this to someone who wasn't a teacher, do you think they would look at it and just go, I have got no idea what they're talking about, whereas actually sometimes you could figure it out.

P3: Yeah, think sometimes that's true. Actually, you see this one, I didn't understand, but I think I saw the root of this problem. But I know we're not meant to look at it, but it's very close to a priori, it's one letter missing. But anyway, we don't say it, That's another 49,

P4: if I understand it...

P3: is a prerequisite for all students. But I didn't know if they meant because they've talked. We suppose we haven't got the whole context here. Maybe it doesn't matter that they talked about reading and writing and was that prior to reading and writing, or prior to what I see? For me, it doesn't. I wonder if it's if it's 54 for me because it's more detailed. Well, it's wrong, I think. But do you think cars in the country are lower? It is a bit like that, actually. Prior to what if they'd said listening, it comes first or something.

P4: Prior to the other skills.

P3: I think it's quite 54ish. Is could that be? Yeah. Okay. I might put a two there.

P4: Yeah. I'm not quite sure about that.

P3: You want a two have a two as well?. We might as well agree on that. Not certain. before you use the new knowledge. So talking about register again? Yeah.

P4: Oh, yeah, 55. You need to hear that's the wrong word, isn't it?

P3: Yeah. I don't even quite know what they mean actually, out of context.

P4: Listen to is it understand what it means will take on or understand.

P3: Yeah. Yeah, I think its that,

P4: I don't think it's listen, is that actually coherence?

P3: I think it is because I think this is different. I just would like to know what they meant. the word exists. Of course, an error with a single word. The word exists in English.. Does it fit into any of these verb incorrect phrases?

P4: But it could be a verb. Yeah. That one, so misselection 46. Yeah. Yeah. Because if it was the whole phrase, I'd say it was incoherent. But if it's just that word, then it's true. It's just it's just the wrong word. (.....) Oh professor. That's spelling. Yeah. Where's the spelling number? So it's over inclusion.

P3: Oh, an extra F.

P4: Yeah, number ten.

P3: That's puzzling me, this. Can you do it with the arrow? Maybe I should be using this. Oh, okay. If it's simpler then I'm doing it.

P4: Oh, yeah. Because I have to....

P3: Professor. Now we're looking at the e misselection, spelling.

P4: Yes. 11.

P3: Yeah, this is one that puzzled us really.

P4: Yeah, and I think...

P3: would you think this one need a bit of extra elaboration like we had before, the priori one, the 54.

P4: I think it's worse because I just don't know what they...I man I assume they mean without speaking.

P3: Because they're talking about the listening skills. I think I do know what they mean, but I still find it rather tricky to categorise the error.. out any voice.

P4: So is it underspecification? But it's wrong.

P3: They're all wrong though, but they are all errors. I think I might do a 54, not 100% committed cos I can't quite see what else... do you think It's more like, you know, you don't know what they mean altogether,

P4: but I do

P3: You do sort of. I think I'm going..

P4: I think is a 54, isn't it? Because there isn't anything else like you say unless it's completely incoherent, which it just isn't.

P3: I'm going to put two though, just to show that we're not sure we've nailed it. And okay, as for speaking, actually that's to me out to be errors. As for speaking, it is an imperative skill for every foreign to own a case, so Ok a wrong verb.

P4: Is it a 46 Yes, incorrect misselection?

P3: What's going on? Is this weird? Who wanna? So is it register because we know what they mean.

P4: I think so. So 55?

P3: I know what you mean about sometimes things are worse though because register wise, that's oddly odd where you found out where you learned to do that.

P4: Whereas in my opinion is not quite it's not as bad as...

P3: it very different qualitatively, isn't it? Yeah. Whatever a student here and write that means no matter what a student hear. Yes. So the wrong word. What is that even classify that as a?

P4: Is it an adverb?

P3: Well, it might be an adverb by the fact that it's none of the others, because that's what adverbs tend to be like, going from becoming multilingual. Yeah.

P4: Well, it's Miss selection anyway, isn't it? Of whatever it is. It's oh, could we have a 19

P3: and then with a single word word exists. Yes, it does. Does not fit into the other B2. Yeah, I think is not a bad idea. Okay. It's an unusual one. Yes. I'm going to put two happy to have 3 or 2?

P4: A 2 two, I think

P3: because that specific whatever a student here, if a student cannot talk. Oh, that's just he doesn't know. that cannot happens to be one word in English that's quite minor really. In terms of the cannot. Yes, but it's spoils the spelling Ones and dinner table and that sort of thing.

P4: Oh, two words. It's a two words Should be one. A4 Yes, it's a 14. That's good. I wouldn't have remembered that.

P3: Talk own views where we want something like Express.

P4: Express. Noun, is that a 27?

P3: 27 misselection. Yep.

P3: It is vain for the student to have so many good skills. Well, that's been taken as an entire phrase, That's the way it's been done. And the whole phrase is a bit incoherent. But we do know what they mean. I suppose

P4: It kind of means worthless, doesn't it? Or pointless. Useless. Or do you think it's a B3 phrase error, miscellaneous, misselection, more than one error in the phrase. problematic phrases which require rewriting. Yeah, that's what it is.

P3: It's 49, it's the whole thing that's a problem. Misselection. Mainly, yeah. Yeah, 49. Speaking to others and get more information from others is a correct way to study academy. That did fox me a bit actually, but then I realised that he meant academically. Do academic study.

P4: So That's all wrong, isn't it?

P3: So it needs a bit of re ordering and wrong derivation

P4: So is that the same thing. Is 49 or is... it Because I didn't just really understand it.

P3: I think I might go for 16 which is the just I don't know what he's talking about. Took me a second. Yeah. Yeah, you might be right, the passing person might not.

P4: Or there is 49. Mm hmm. Yes, 49.

P3: Did you say 16 words do not exist in English

P4: or meaning of whole phrase is unclear?

P3: Yeah. Maybe 16 with not 100% confidence. Maybe that is closer. I don't I think someone reading it first wouldn't get it. I think I got it once I analysed it little bit.

P4: Or do you want a 49 which would be meaning clear, maybe more than one error? Yeah. Phrase requires re- writing. Well, yeah,

P3: Probably 49 actually.

P3: So you're 49? I'm going to have a 16.

P3: Okay. I think that's 49, 2, 16, 2. Yes. I think the most important is not only reading skill, that's just not. Yeah. Taken as a whole, it doesn't make really proper sense

P4: because you've got the most as well. You can't say reading is not only the most important skill, it doesn't kind of.. The most and the only... can I have a 16 please? I kind of get it but I just think, well yeah

P3: I might do the same, actually. I'll do a slight 2 for this one.

P4: I'll have a 3 just to be.... Yeah.

P3: I think the most important is not only reading skills, other skills are not either. That to me was did disintegrate. I felt that was unclear.

P4: Is that the same or is that a 54?

P3: I cannot be sure of the meaning of the word effect. I would call that 16 actually. What's it when you say..... Improve...the expression is unclear. Categorised as B1, a little bit foxed by

P4: or 49? it's not, that is it,

P3: Omission is not allowed. So there's some stuff missing. You see

P4: other skills are not either. I feel 16, I think if the phrase before it had made sense. Reading is so say it said reading is not the most important skills. It's still not good is the other skills are not either. It's the either. Okay. Yeah, 16, please. Okay. I obtained the biggest. That's a funny...kind of all of it, isn't it? Obtain the biggest. Because Obtain the biggest?

P3: what we know what they mean to make progress.

P4: Yeah. The most appropriate. Yes.

P3: Suggest maximise progress. Yeah. So we know what it is meant but the phrase is, is wrong.

P4: So is this a 49 or is this a 50

P3: meaning is clear, but word choice is a non native, right?

P4: Yeah. Yeah.

P3: Yeah, I think Happy with 49, Yeah. Okay. Yeah. I think we finished, Tony.

Appendix 6.9.3 P5P6 LEA transcript

P5: to get a perfect academic performance is a basic quality that every student wants.....

P6: So what's this wrong word? Sense relations.

P5: Yeah, I was thinking that, but I was also looking at collocation...Arbitrary combinations. They are fixed phrases that was connected down to B8, but it's not B8.

P6: I wouldn't say to fixed. I think it's like this is wrong sense.

P5: Cohesion. Lexical sub phrases. Lexical subs referring to the same subject, Incorrect choice A. Yeah, I agree. I think B3. Which is misselection. Right. Which is number 19. I'm going to go I'm going to put it as three. Confidence

P6: Same thing. Yeah. Well, this is why I think most of our scores are going to be the same discussing it first.

P5: Okay. Second one gets spelling. Yeah.

P6: A3? Spelling three.

P5: Yeah. Which one though? Is it omission because he's missing a couple of words?

P6: It's omission. Yeah. A31.

P5: That means it's number nine?

P6: you're correct. It's number nine.

P5: I'm going to put myself as being a confident person. How about you?

P6: myself? That's the same way. Perfect. Basic quality.

P5: Basic quality..... I'm tempted to go for two words should be one here.

P6: Two words should be one? Basic quality should be just quality. So you think that's an over inclusion?

P5: Yeah,

P6: I was going to say why do they consider that to be an error anyway, The word basic. I like the idea that

P5: Yeah, yeah, I know what you mean. 14. I'm going to go for it.

P6: Yeah. Run me down as a 2.

P5: Me too. Let me discuss let me discuss the topic.

P6: What about B12?

P5: Yeah. Yeah.

P6: Overly informally? In this sense.

P5: Yeah. You're right. Yeah, 42?

P6: Yeah.

P5: And I indicate how And I indicate how I disagree with the topic. Indicate how I disagree with the topic. What's the error here? Is that an error? Disagree with the topic?

P6: Well, you wouldn't disagree with the topic because you disagree with an argument. You disagree with the stands the point of view.

P5: So you suggesting it's the wrong word?

P6: Yeah, I think it's the wrong word.

P5: Okay. But it's it's got a relationship. So it's a number. 19? Okay. I'm going to put myself as less confident though with that one. You put number 2? Reading skill undoubtedly is a very semester. I mean, skill.

P6: Yeah. You can see use this one sentence for three because you think the whole thing is shocking.

P5: I would again say this is a read two words should be one because you don't need skill in there. You just read is a skill, isn't it? It's a very crucial factor to me. Yeah, So I'm going to go through that. Yeah. So two words should be one, which is number 14. I'm going to be very confident you look the same. Reading skill undoubtedly. Okay. So spelling, and that is, there's an omission, there. Let's just go back to that one, so reading skill could be..... what do you think?

P6: Well, it's not connotative error, meaning it's no, it's not verbosity...

P5: is it cohesion? Is it like cohesion?

P6: It seems more like cohesion to me than anything else

P5: over inclusion. Maybe 24.

P6: Yeah,

P5: I'm going to drop my confidence level. You happy with 24?

P6: I'm happy with one as well. Okay. That's not the first time that we use that problem.

P5: No, we used it somewhere else here. Basic quality. Yeah. So it's the same idea. Yeah. Because we found is a misselection may be of 24, a misselection of the word. I'm happy with that.. okay. So here's a spelling if you agree with that. And that's going to be omission Number nine, is it? Omission three?

P6: Yes.

P5: As reading skill undoubtedly is a very crucial factor to master.

P6: Again, misselection of word

P5: watcha think? Wrong word sense relation.

P6: Yeah, yeah,

P5: And number 19. Okay? myself. Confident, happy with confidence as well?

P6: Yeah.

P5: As a student who finishes amount of tasks and homework,

P6: this is stylistic. It should be complete.

P5: So is that the wrong word? A sense relation error again?

P6: Yeah,

P5: Do you think? I think it is, yeah. Rather than a connotative meaning. Is it a bit odd? Connotation,

P6: Really odd.

P5: Is it? no no go wrong word. Which we'll do 19 again then?

P6: Yeah. Yeah.

P5: Okay. I must use lots of knowledge because I do not know how to make it better. I'd say this formality.

P6: Yeah.

P5: Because it works grammatically. I think it does.

P6: 41?

P5: A miss selection? 41.

P6: Yeah,

P5: documents that I do not know to make it better, Documents, actually documents, the whole thing. Document that I do not know. Documents that I do not know.

P6: And this one's a little bit more serious. I don't even know what that word means. words in same lexical set and in the intended word.

P5: Okay. Would this be cohesion? Because there's a need for kind of maybe like pronoun use like a discourse,

P6: you just dump it under B5?

P5: Yeah, that I do not know To me. You get rid of that, I think.

P6: I don't know what about B6

P5: could simply be nominated like (...). Yeah, that's true. It's clear, yeah.

P6: But could be said in a better way?

P5: Written in a better way? That's a good one. Should we go for that. 27? good find. Are you going to put three confident?

P6: Yeah

P5: I must use lots of knowledge and documents that I do not know to make it better. That should be improved, Right? Wrong word sense relation error again, do you think?

P6: Yeah, we're being like students now. We're beginning to wonder why we're using the same number of

P5: Yeah. Yeah. Yeah. To be something wrong. Different kinds of books. Essays. Right. So that's spelling misselection of spelling. Is it a blend? They've missed some letters. They've added a few letters. I'm going to go for blend personally.

P6: What's blend?

P5: It's two items are incorrectly combined in some way.

P6: Blend. This? Yeah. At the bottom.

P5: Are you with me?

P6: Yeah. He's got the ending of Y.

P5: He's missed a Y and he's added an I.

P6: Yes.

P5: And so on. Different kind of books, essays, files, and so on.

P6: Again, formality 41. Hold on, Misselection?

P5: Yeah. In the end of a semester proposition, right?

P6: Mm hm. So what was the particular field?

P5: misselection? Yeah, 22. Is it the most important? I don't think so. For me that's formality because I wouldn't put a question in an essay like that. Agreed?

P6: Yeah.

P5: Yeah, 41.

P6: We actually teach them not to do that. Yeah, I don't think I don't think so. As opposed to I disagree. again. It's getting into your formality.

P5: Yeah. So I guess the question of formality versus it being slightly odd in mean. No, the meaning clear isn't it? words are Okay, So it's formality. You think again?

P6: Yeah, Yeah.

P5: So what score a student can obtain is determined by how to organise the words. How to organise. Paraphrase?,

P6: I would go for that.

P5: Yeah. Yeah, Happy with that?. Yeah, (...) in my opinion. It looks fine. Dramatically, right? So what's the error you think? Well, the use of personal pronoun informality.

P6: Or overly formal. Or overly informal. Yeah. In the academic essays. informal.

P5: Yeah. Shall we put that then?

P6: Yeah.

P5: The most direct approach. So I suppose the word direct could be effective..... I'm gonna go for direct to the wrong word.

P6: But is it? I'm gonna go for..., no, It's not paraphrase. So it's not cohesion. So I will go with you and put me down as a (...).

P5: To gain great academic performance.

P6: Well, just a clumsy way of saying it.

P5: Yeah, that feels like B six to me. The paraphrase one. Yeah. What do you think? It kind of makes sense, doesn't it? Something a non-native speaker would say but just needs rewording. I'm for 27. Listening skill is a prior skill for student,

P6: You wouldn't say this speaker obviously incorrect word choice, it doesn't collocate.

P5: No. Is there a field error there then? Do you think you think it's totally wrong or is it is it relatively clear? we've not gone for that field error , it's totally wrong. Shall we go for 26? Yeah. Sorry. Oh no. 18 field error or should we just say it's a close, prior skill? I don't know what the meanings there. Like essential skill or is like a pre existing skill.

P6: Yes. So we said it's not

P5: I'm going to go for 18 thing is that it's it's completely wrong so rather 19,

P6: Missselection.

P5: All right. Yeah. How about you? Same? yeah. Before you use new knowledge, you need to hear the content of what professor says clearly.

P6: again, Formality.

P5: Formality, Yeah. 41.

P6: Yeah. Yeah.

P5: Before you use the new, you need to hear the content.

P6: Well, again, this is an emission because it says you need to hear content as it stands. It's not a word choice or anything like that because it's grammatically incorrect. No.

P5: Or should it be like you need to understand the content of, you know what I mean? Like the verb here, you said the wrong word.

P6: So how do you cope with the fact that they've got two errors? Yeah, try because even if you wanted to use understand afterwards.

P5: Yeah, but yeah. I know what you mean. But if these are just lexical errors, does that make sense? Yeah.

P6: Well then are you going to call it paraphrase problem non-native-like phrase?

P5: I'd be tempted to say wrong word again? Actually. Tempt to say number 19 again.

P6: Okay.

P5: Put it as two because we're not... content of what professor says clearly, so spelling and it's a misselection I'd say, Right? Yeah.11.

P6: Where's the where's the (...)?

P5: No problem. Yeah, so that oh, just the ten. That means over inclusion.

P6: Over inclusion,

P5: which is say ten, isn't it, yeah. So we'll go back to that one

P6: and that is miss election.

P5: Should be a no. Should be.

P6: So we're right on that.

P5: Yeah. Have any voice? Paraphrase?

P6: Yeah, it's still native-like

P5: Yeah.

P6: 27, 3 27, 3

P5: As for speaking, it is imperative skill for every foreign student. So it should be have. So again, that's got to be wrong word in the right area. So it's 19. Who want to be a silent person for formality, right?

P6: Yeah,

P5: yeah, Whatever a student hear and write. And (...), so. All right. Is it formality, whatever, I mean, I'd say it wouldn't be whatever a student hears. The word whatever.

P6: It's not an academic word

P5: formality, that one.

P6: Miscellaneous

P5: what? Like miss selection. Just the wrong genre or the wrong,

P6: what alternative would you put there?

P5: like, you know, irrelevant, off the topic. (...) And it's a little bit of a mess, isn't it? That is a mess?

P6: I think it's a mess, so.

P5: We could put it as miscellaneous and over inclusion like it shouldn't be there.

P6: Yeah, because it actually, you're right, It's not the right word at all. Should say it should be irrespective of what.

P5: Yeah, Yeah,

P6: What do we want to change it to?

P5: You see, I wouldn't put that. I would say it's formality issue. Spoken rather than written discourse.

P6: Oh, in that sense, yeah we could do.

P5: Should put it's a 41 do you think?

P6: Yeah,

P5: leave it at two confidence?

P6: Yeah

P5: If a student cannot talk own views. If a student cannot talk own views

P6: Does somebody believe that cannot is one word?

P5: Ah, yeah. Good man. Well done. Two words should be one. We got there, we finally used it.

P6: incorrect. Selection.

P5: Yeah. Use this a lot.

P6: Yeah. Yeah,

P5: it is vain. Oh, it's all, this is vain. Have so many good skills.

P6: Again, paraphrasing, there's more than one problematic phrase which requires re-reading.

P5: Yeah, I agree.

P6: We used that a lot as well.

P5: Way to study, to study academically.... Oh, it's going to be, suffix should be academically. How about suffix type?

P6: Why wasn't just academy highlighted? Then?

P5: I guess to kind of emphasised the fact that it needs to be a verb with a verb. So I'm going to say that is an omission of a suffix. So number one

P6: or is it a missselection? No, it's ommision.

P5: Yeah. Yeah. I think the most important is not only reading skills important, there's a missing word there for me. Which I feel like, oh yeah, there you go missing word number 40, Missing word happy?

P6: Yeah.

P5: Other skills are not either. Gosh, I mean, that's a B11.

P6: I'm looking at the Bs., I was oh no, I was looking further up.

P5: Or even B5 cohesion. How it's not really stylistic actually.

P6: I would have paraphrase again.

P5: They should be connected, shouldn't they? So B11 is stylistic.

P6: yeah, I was looking at B11.

P5: Okay. Go for that? The best way to obtain the biggest to obtain the biggest progress looks like a collocation. Obtain.

P6: Yeah. Well, not just collocation but also this is an unnatural way. Yeah. You don't obtain progress and you don't have the biggest progress. So both of them are the same.

P5: Do you think paraphrase or wrong word.

P6: I go for paraphrase.

P5: Would you? Okay.

P6: Number two.

P5: Yeah, me too.

Appendix 6.9.4 Post LEA interview P2

Interviewer: Interview with Participant 2 Hello. Thank you for coming. Just a few questions if you don't mind. First of all, may I ask you about your number of years of teaching experience, how long you've been teaching?

P2: Yeah, I've been teaching for ten years.

Interviewer: Ten years. Okay. And that's great. Before and during this career, can I ask you about your linguistics training, the study of the language itself, What have you done there?

P2: We focused on linguistics during the Delta, the diploma in English Language teaching, and I've read for the MA, Applied Linguistics and TESOL. And so both of these areas, there was more focus on teaching English, second language, rather than linguistics.

Interviewer: Okay. And have you done any other reading outside these courses, into the language itself?

P2: Yeah, I think it comes up for most teachers, even to the point of just into language textbooks. You have to have a language awareness in order to answer students' questions. I guess most days there's some linguistic awareness that you need to read around.

Interviewer: Okay, that's great. And so I'd like to ask you a few questions now about the document that I sent to you. I don't know how much time you had to read the guidance which accompanied the framework, the guidance on how to identify separate lexical from grammatical and the categories themselves. Was it clear to you when you read it?

P2: Yes, it was quite clear. The separating the lexis and lexical areas and the grammatical areas was difficult and the guidance helped. But there were still instances that I felt that that could be grammatical, that could be lexical. And especially when you think of examples, I guess we're used to, when we provide correction codes or error analysis, we're used to not separating them. And those are, two are so intertwined. So it was a little bit confusing. Separate, reading the instructions about how to separate them. But the example certainly helped.

Interviewer: Thank you. Okay, question 4, you've already answered that you've already explained. To answer question five, the categories in the framework now, and are they clear? Are the categories, the types of errors to be allocated to them sufficiently clear for you and distinct from each other?

P2: Most of them were. I was glad to have the framework before we used it to get my head around A few of the categories such as B8 arbitrary combinations and irreversible binomials. Just having the additional supporting document explaining what that is and giving some examples was useful. Things like coherence, words do not exist in English or the whole phrase is unclear. Cohesion, lexical substitution pronouns, used and conjunctions. Sometimes I find it difficult to differentiate between those two. It was generally fairly simple to decide which vertical column to use. The omission overinclusion, misselection and disorder.

There was an explanation of what those meant, but it was pretty clear what overinclusion means. But I'd say perhaps going back to the, the guidelines for the framework, some of the terminology there, derivational affix error, inflectional affixes, genitive errors. You obviously, we know what those mean, but I don't know if every TESOL teacher would. I think there was certainly one I had to Google beforehand. And it does take you back to your kind of training because we don't use these terms, certainly not with students because I think they'd be scared of the language, meta language. There was times that I had to, I had to get my head around it before knowing what it was. Certainly, when we were using the framework, it became a little bit simple towards the end as we've had some practise with it.

Interviewer: Thank you. Question five. So, sorry I didn't ask you specifically for question five. Would you say completely clear? Quite clear? Quite clear? I thought so. Okay. Question six, then you've already answered. question seven, but the differences between the categories was clear enough for you kind of already. Sorry, it's kind of already done that one too. Yeah, the focus here though, is in question seven is on how easy it was. I'm particularly interested in hearing about instances, if you can remember, any of an error which perhaps couldn't be categorised or an error which perhaps could be categorised in more than one area.

P2: Yeah, that happened occasionally and I think performing the task with a colleague helped identify that. If I said, no, I think you should go here and she said no, what about here? You have to discuss it more, and then thought about the other person's point of view and their categorization, and that could be true as well. Let me just think about more. That's strictly true. Um, , sometimes when an error then changes the next word if it's an error with plural or singular and that change changes the following verb, sometimes you think, well, that's an error because of maybe lexical error. And then where does that fit into it? And then the next one is a grammatical error. When you fix the error, it then makes another one, which categorization does that then fall into? That was quite difficult. It helped Having the table with the italics word. That's the word you need to focus on for this specific error. But there was still some.. such as before you used the new knowledge, we both eventually agreed it was 41. Formality. That was a confident, sorry, I'll find one that we're not confident about. I remember there was some discussion for the first one. Okay, Get to pose, that's academic. Okay. Yeah. Because we agreed that was eventually formality. But it does depend on the text the student's writing, because it could be to get a perfect academic performance. To achieve, achieve an academic performance. We weren't sure. Sometimes you're not actually sure what the right word is or what the error could be resolved to be without actually discussing it with the, with the student.

Interviewer: What we're talking about here is a plausible reformulation of what they meant, which is sometimes perhaps required in order to you to identify what the problem was with what they said. I think that's sometimes the case myself, but perhaps not always. Sometimes we can see, well, that's just wrong. I don't know what they meant, but I can see that's wrong. That's the type of error that they

made. But as you say is plausible reconstruction is important to get into their heads. You can always have time to meet the student to find out.

P2: what I think about going off track when I look through student's work. So as you said, very simple to identify the wrong word Form suffix is incorrect or the preposition is wrong. And then others I usually write, what do you mean here? But then that's because I can I afford the time to sit down with students and say they put it in their words? What did you mean by that sentence? Using your own words and then you can understand that's what they're getting at., but that's not always possible. I think using this is time consuming, but not as time consuming as actually meeting the student to find out what they wanted to say.

Interviewer: Okay, thank you. We have now. I just want to make sure we've got answers to these quantitative questions. Maybe you could go through them just so .. overall, How easy or difficult was it for you to use the framework?

P2: Because it was my first time using it. I'd say a little difficult. But that certainly became towards the Quite easy, towards the end. I guess with practise it becomes easier with familiarity. and I think it would certainly be easier if I was using it alone because then you wouldn't have to rationalise all your choices all the time. It certainly helps to verbalise what you're thinking. But having another person to literally second guess to you. Right. Okay. Sometimes I know for myself when I mark things I can be quite quick to jump to. That's what the problem is. Maybe I do need to slow down really thick, ponder over. And this forces you to do that. Could say wrong, but why is it the wrong word? Is it field error or sense relations? Actually, those two. At the very beginning of the task, I found it was quite difficult to differentiate by the end. It was really quite simple. Okay. That's about the general, that's about the specific, whatever it was. Yeah. maybe I should go for not sure in the middle because it started difficult and ended easy.

Interviewer: xxxxx, thank you. Now, I'd like to ask you about the depth of analysis, the results that something like this would throw up. I'm interested to hear your thoughts about, do you consider this an exhaustive taxonomy of the type of errors which students might make? Do you think this would constitute good feedback? What are your thoughts?

P2: Some of them are certainly used more than others and some of them could be lumped together. Just going first to suffix prefix, there are two separate categories, and yes, they are two separate categories. But for perhaps simplification, you can put them together. Suffix /prefix because probably a student is not going to make an error with both at once. Having an error with a suffix and an error with a prefix, you just write it twice. Some of them perhaps could be combined. What else? Some of them we didn't use so much. Perhaps miscellaneous, wasn't there? There's a temptation with miscellaneous, we really try to avoid it because otherwise you're writing everything's miscellaneous. You're writing everything. I'm not too sure which one that is. So it's I'll just put it there. Same with paraphrase. Really. Paraphrase, I'd probably only use if it, if you're using an academic text, you say, write it in your own word because otherwise it's, paraphrase's a lot of problems here. That's perhaps when we tried not to use, during this framework, we

wanted to focus on what the others were. Do we sometimes have a problem with maybe.... one where we disagreed.. Was it is important to number 17? Is it the most important? I do not think so. My colleague said that was a connotative meaning, whilst I said it was formality. Because we understood that this is an academic text, I think I'd still go for the same. But perhaps because formality is a bit easier to understand than connotative meaning. The meaning is clear but also odd. What makes it odd? For me, it made it odd that it was not in the right register. Perhaps for my colleague, it was odd because it didn't quite fit in, but not because of the register. So that B9 perhaps caused us some problems because odd to whom maybe. But I think it is certainly more comprehensive and detailed than the standard correction code we tend to use, just like WW or WF students in my experience sometimes are frustrated by the correction code anyway. But I don't know what the right word is, I don't know what the right word is. This might help because they have to think about why is it the wrong word? Is it a field error, sense error? Or it might confuse the students, like what does the field error mean? I think it certainly should come with an explanation for teaching, use of. You need to explain the benefit of using students. Obviously we know the benefits of Error codes, but I always explain why I've written random symbols and letters all over their work and how that will actually help their language development, because otherwise they don't see that initially.

Interviewer: Yeah, Okay, thanks. We've already got into the next question. How might these results be used? Now there's a variety of ways you could look at one error, one essay from a student and copy and paste these errors into these boxes if this were a bigger document. Or you could tally them where you could get a, a set of essays from the whole class and then tally the errors into these boxes. Is there any of these approaches of interest to you? Do you think you might consider using them or is specific lexical errors perhaps not something that you might want to focus on with your students?

P2: I think it depends if there's something that's coming up quite a lot. I don't think I go through the stage of, every time I have a bunch of essays filling out this separate table of where the errors were. But if there was something that I think, okay, a lot of students are making the same errors with, I think, prepositions. It might just raise your awareness rather than a lot are making the issues with wrong words. You've probably just dismissed that. Well, not helpful. Whereas, it draws your attention into certain areas with formality and then you could perhaps do an exercise task or a lesson on those areas. I think that the formal lexical errors, I think it'll be harder to do because maybe you pay more, a bit more attention to spelling or word combinations in class, but you wouldn't necessarily then do a whole lesson on it. So I think it arises there, there was an error and there was a common error in the, in the group of students.

Interviewer: Okay? So it might give some opportunities for remedial teaching in some areas, but not all. Obviously, as you said, you couldn't say to them. Oh, yes, a lot of you are making mistakes with the wrong word. Do be careful with the wrong word. That's no help to anyone. But I think perhaps.... Okay. No, that's fine.

That's great. Thank you. Give any other comments about the guidance or the framework for improving them or any other comments in general?

P2: I think you'd have to have some training for teachers to use it, especially those that don't have, I suppose, a strong linguistic awareness of meta language because it's quite meta language heavy. To make it a little bit more accessible and teach a very, as I said before, the derivational affix errors. Perhaps more examples would be really useful. Because I found that the best thing to get ahead around some of the differences was, oh, that's an example of that because we're so used to using wrong word or word form or the general terms for these examples, we're more used to seeing examples rather than these new categories. I don't know if it could come with the needs, Some training or like a little self test to make sure you have you got it right Quiz that they could do before they start to use it right. Because it might be then if they're not using, the teachers are struggling to use it, then they go to the standard miscellaneous or then revert back to the whichever method they've been using before. And the separating the lexis and the grammar. I think I think you realise that that's nearly impossible, it's really difficult to do. I suppose sometimes I have written in my correction code, I've written a G for grammar, which is quite basic, but it's usually for if they miss the plural s or the third person s, It's clearly just a little grammar thing, you know, there's certainly mistakes rather than errors. It needs training guidance for it.

Interviewer: for sure. Yeah Okay. Well, thank you very much.

P2: Thank you.

Appendix 6.9.5 Post LEA interview P3

Interviewer: Thank you for coming and thank you for agreeing to speak with me. I've got a few quick questions here. Number three, for example, how clear to you was the guidance, the guidance which comprises this and this really the guidance to use the framework.

P3: Quite clear.

Interviewer: Quite clear. Okay. Thank you. I'll ask you why shortly. Are the categories in the framework, are the categories, these different categories in the framework sufficiently clear for you, distinct from one another?

P3: Yes, I thought there was good distinctiveness even though occasionally there would be different ways you could do it, one knows as different ways of organising these things. But I thought, again, pretty clear and very helpful. So that would be quite clear.

Interviewer: Quite clear. Good. Thank you. Question seven, were the differences between the categories clear enough for you to use the framework? Obviously, I've recorded your conversation with the other participants, so I'm aware of where there was disagreement between themselves and between other participants. But can you remember any specific issues with a couple of categories being quite similarly?

P3: I can remember that there were a couple of occasions where we just quickly needed to sort out what was meant by the categorization, but I don't think it was ever to the extent of confusing or particularly conflicted. It didn't stop us moving forward. So I think again, I go for quite clear.

Interviewer: Okay. Okay. Thank you. There we are and the last Likert scale question overall, how easy or difficult was it for you to use this framework?

P3: Yes, there was several occasions where we had a mild hesitation, but I did find it overall quite useful. So I'm going to say quite easy.

Interviewer: Okay. Thank you very much. So, just a few qualitative questions. Can I ask you how long you've been teaching English as a foreign or second language?

P3: Over 30 years.

Interviewer: Wow, that's great. And what training in linguistics have you had?

P3: Linguistics, MA level. General linguistics.

Interviewer: Uh huh. Have you done any other reading or courses about linguistics in general?

P3: Yes, quite a lot of CPD. And also I find that the kind of explanations attached to course books are very practical as well. And I also occasionally read more scholarly articles, although it's not particularly my area now.

Interviewer: Okay. So you said that the guidance was quite clear and could you say why you said that? And any tips you can give me for improving it maybe or anything which is hard to understand or

P3: there are some categories where you are taking a while to think what does that actually mean? I'll just give one example, something like connotative meaning. That would be the B9 one. I had to think what was meant in this situation about connotations. When you get an irregularity, like he's quite notorious for the charity work, you could say for instance that notorious is a gradable adjective. You could call that, for example, a gradability mistake. Like it's very expensive or it's absolutely extortionate. You know, we grade some adjectives, we don't grade others. So there were moments where I thought, well, you can look at another way. However, I still think it's useful to have some category. Even if it's one that you might have done another way. I don't think there's ever going to be a perfect way to distinguish every single aspect of this. That's not what language is like, really.

Interviewer: Mm hm. Okay, so you've given one for question Five, are the categories in the framework clear, the types of errors to be allocated to them sufficiently clear?

P3: Yes, I don't think I would ever say. I think generally there it was always a help to have a categorization. My only question would be that the categorizations themselves could be seen in slightly different ways. But it's very useful to have a way, an organising principle, and I think that's really useful. another thing might be an arbitrary combination. Sometimes something looks on the surface to be arbitrary, but actually there's an underlying rule in a chomskian sense that someone is following, that they've taken from, they've overgeneralized from somewhere else. How arbitrary is it? It's usually not really arbitrary or random when someone makes a mistake, as you know. Again, that's something that you could as it will have a philosophical discussion about, but it's still handy to have something to call it. Very practical, I think,

Interviewer: indeed. Thank you. Yes. Are the categories clear? The differences between the categories clear enough? I think we're repeating ourselves. You've mentioned quite clear. In answer to that, you've already mentioned connotative meaning Could be also a gradability adjective error. Are there any other areas which you think has overlap or any Omission of any kind of area there, do you think?

P3: I don't think you'd want to have any extra categories due to the fact that for practicality purposes, you can overload the user if we're talking about practitioners using them. You know, you're always going to have B13 miscellaneous, aren't you? I don't think perfect refinement is possible. So for that reason, I've not said it's perfect, I don't think that's possible. On the other hand, I think it's really quite helpful. It gets you a long way further and we get the average non linguist practitioner a very long way. You know, I think I'd call it good enough, personally, if that's like the sense of the good enough parent, you know, good enough is a threshold category which makes it actually a very, very useful piece of work.

Interviewer: Okay, thank you. Thanks for that. Overall, the ease or difficulty you rated as quite easy was that because of the possible perennial problem with dual categorization?

P3: That's right. Just a slight hesitation on interpretation. It was only slight we found.

Interviewer: Okay. Any other kind of issue with either the guidance for description of how to identify errors or how to categorise them? The framework itself?

P3: I actually thought it was well thought through. Actually, sometimes things were bundled together. For instance, number six, the errors in countability, tense, grammatical redundancy. My mother, she is ill. That's just the grammatical redundancy. And that's quite a bundle, isn't it? Whereas others are very specific. I mean tense, that's pretty huge, isn't it? For Someone to make a tense mistake in English language. It's a real problem, isn't it? That could be something where I'd hesitate if I was making one.. I might privilege tense quite high, because you're going to hear a tense mistake often. With some speakers, I mean, many languages don't have tenses as you know.

Interviewer: Mm hm. Okay. That's great. Thank you. So this analysis, it produces a certain depth of analysis and the results that this activity throws up. Would you be satisfied as a learner or a teacher, that you could provide what might be described as a good depth of analysis to students? Or do you think it's perhaps quite a surface level analysis at the moment? Is there too much information to give back to the students with the resulting analysis? Is it good for the teacher? What are your thoughts on that?

P3: Right, my thoughts are pretty confident on this. It's good to have this level of depth because it almost acts as a teacher training vehicle. This would be developmental for many practitioners to use this. And that's a real positive. That doesn't mean when I train teachers, I will always say this. That doesn't mean that you offload this level of depth straight onto your feedback because it won't make sense to many learners. In fact, it would be too much depth for the average learner in my view. But to clarify for the marker, what's gone wrong is very important. And then there would be, I think, a training need with some practitioners to know exactly how much that is transferred onto the student. The student doesn't necessarily need to hear phrases like what you've done There is a collocational error relating to a fixed phrase. You would not say that the student to help them, but you would say simply it's, it's not hike hitch, it's hitch hike. And you probably teach that completely. Behavioristicly, make the correction, you don't need to call it anything.

Interviewer: Bi nominal, irreversible, bi nominal.

P3: I mean, that would really not help anybody. You'd be distinguishing between the usefulness for the teacher and the usefulness for the student. There might be students that will then ask you, why is it wrong? Well, if you want to know you've done a switch. I'd still probably do something like this with my finger, like I might

the tense it had gone before. The simple past, so used with discretion by teachers. It will be pretty powerful.

Interviewer: Okay. Thank you very much. How might these results be used though? Let's say the results for the student, could they be used as if they were to submit one student, one essay for each student, which is a lot of work, isn't it? Or maybe there's another way. Maybe that the teacher could perform a lexical error analysis on a body of student work and present the overall results to the class. Or somewhere in between, have you got any suggestions or ideas about how this might have any practical implication?

P3: Well, I like your suggestion of doing some class work with recurring errors. I don't think you can actually focus on every single thing everybody does wrong. That's not practical. It's too much work for everybody. But I think a teacher would probably, using this, be able to quite swiftly spot error patterns amongst the group. Possibly if you do lots of one to one teaching, it would be very useful to work with that individual and lots of tesol teachers do one to one. Actually, I think they would be quite grateful to have a systematising scheme like this to work with somebody. For example, for many years I taught individual businessmen and you know, you taught them all day one to one. And is my grammar okay? What's my writing like? something like is very useful to the basis of the discussion. The thing that you're doing quite a lot of here is this tense error or actually it's more collocational, You're very good. But these collocations will just learn a few each week. I think it will be useful, but in a large class, you'd look for the bigger patterns and in one to one, you look at the pattern that the mistakes that relate most commonly to that person. But again, I think discretion and professional judgement are going to be quite important. You wouldn't want to teach loading back all this onto the learners. And would it be valuable? Would it just make people think, I hate grammar. I, I'm not good enough. It could look like the deficit was more extreme than it was. You want to tell people where they communicated well as well? Yeah, I think probably a nicer guide for the teacher would be a good accompanying handbook document for this.

Interviewer: Mm hmm. That's a good suggestion. Thank you. So that answers number 13. Do you have any comments on the guidance or framework? Some kind of handbook, maybe in a more reader user friendly way? Exactly. Something attractive with some illustrations. just in colour, you know, because otherwise this could be heavy. People that are not interested in the linguistics. I'm really interested, I enjoyed it. But not every, if you're thinking about the general tesol, teaching world, I'm thinking about people like my son who's a professional. So 30 year old, he might ask me what's an adverb Again, he's really, he would be overwhelmed by this. I'm not saying he's typical, but he's not necessarily unusual in the bigger picture, is he?

Interviewer: Certainly there's some terminology here which threw a few of the current participants.

P3: Well, you know, and you know, there's different terminologies to explain the same thing as well. That's not surprising. Yes, this is quite a good choice of

terminology actually. But, you know, as I say, you could bring in gradability. It's a very big error as well. You can, you can bring in what you wanted. There are many descriptions of the language to draw,

Interviewer: indeed. Well, that was the last question. Any suggestions for improving the guidance or framework? Can you think of anything else?

P3: Yeah, user friendly and well thought through teacher support material with examples, with comments about how deeply to go into it. With suggestions for different types of students that might get different types of benefits. be quite handy. How deeply to use the linguistic terminal meta language for one thing, I think a lot of students of if they're adults, they're past the stage of picking up grammar in a language acquisition way. And they have a certain knowledge of grammar that they are hoping to activate and to improve their writing. But it's a matter of judgement how much of this maps onto their terminology. What they're expecting to be able to use. Again, careful conversations between learner and teacher as to what is useful and it's going to vary. So that would be in the guidance just just like health warning about not throwing it all back on them to make sense of it.

Interviewer: Yes. Good points. Well, thank you very much.

Appendix 6.9.6 Post LEA interview P5

Interviewer: Participant five. Recording on 6 November 2019. Thank you very much for giving your time. Can I ask you first of all to circle some of these multiple choice response things? And then I'll just ask you the other questions when you're ready.Thank you. Okay, great. So thank you. Okay. So first of all, how long have you been teaching?

P5: I've been teaching for about 15 years now.

Interviewer: Wow. Yeah. Pretty good.

P5: Yeah, 15 to 16 years.

Interviewer: And is that 15 years post CELTA?

P5: Yeah. Yeah. Oh, no, about 10/11 years post CELTA Because my first teaching post is without CELTA.

Interviewer: All right. Yeah. Okay. So I would say that puts you in the highly experienced Bracket. but what about specific training in linguistics? Yeah, as in terms of study of language. And I know it was part of the Celta, you did have to do a language awareness component.

P5: Very true, Yes. I've done the CELTA I've done the diploma, the Delta, I've done an MA in TESOL and Applied Linguistics. I guess it's always been there in my teaching career.

Interviewer: So we could call you not only highly-experienced, but also highly qualified. Okay, jolly good. Question five then, because you've said it completely clear, the guidance, when I say the guidance check with you is this advice for identification? And separation of lexical and grammatical errors, but also perhaps the explanation of these categories and how to categorise. What I'm looking to doing is improving this. I know it was a couple of weeks ago and cast your mind back, I've already got the recording so I know where you categorization will differ from your partner's and how it differs at times from other participants. So I'm not, I'm not specifically interested in hearing. You recall where you struggle to categorise something confidently. What I'm interested in hearing, is there any ambiguity in this guidance?

P5: Well, that's interesting because I was only going for a complete play with these instructions. So if I were to go, I would probably put it quite clear actually because Yeah. Because that was more. The reason being, I suppose on the surface level it was quite heavy because there's a lot of meta language there. And certainly with the elements of the socio, elements, I suppose the, maybe I'm jumping ahead, but issues such as connotative meaning and formality, the scope for ambiguity between the two, maybe. Yeah, but I might be jumping ahead there. Generally speaking, it was okay. But there are elements that I felt were quite heavy and difficult to interpret and discriminate between, I suppose.

Interviewer: Yes. Thank you. Thank you. The question five, the using the types of error to be allocated to them are sufficiently clear. The description of the I just mentioned there was some scope for perhaps dual categorization. One, go into this or that. You got other thoughts that?

P5: that was the main reason really why I put it quite clear. Could there be more examples? I think possibly. I remember when I was doing the exercise thinking this could potentially fall into more than one category. I think I might be jumping ahead again in terms of suggestions, but L1 issues as well L1 one errors. But maybe that could be for another time in terms of categorising an error. But my main kind of query, when I was doing it, was thinking, could this error be more than one issue for the marker, You know, kind of a bit of a subjective question. Is this a connotative error, a B9? Or is it a formality issue they got wrong here? Or is it just a simple case of paraphrasing has gone wrong so that it was more socio rather than that. Again, surface level kind of evident grammar mistakes. As a teacher, I'm much more comfortable and used to identifying.

Interviewer: Right. yes, yes. Well, that is actually an issue with error analysis in itself, the possibility of dual categorization. Possibility of being able to pop an error into one or more of these. This is why this hasn't been done before. That's one reason why if I can unpick this, it might be good. Anyway, thank you. You do that. Number seven, were the differences between the categories clear enough for you to use the framework easily? Well, I suppose we've already just talked about that too.

P5: Yeah. And I was also kind of aware, I don't know if these columns as well. Omission over inclusion, misselection, misorder and blend. That's one of the reasons I put quite clear here is in terms of differences. Yeah, as I mentioned, it was quite heavy on the meta language. It took me quite a bit of time to work out where I should put the error in addition to that possible grey area. Does it fall into which category, and that might be my own lack of experience in analysing errors to this depth rather than my everyday correction code that I'll use with my learners at the moment, which would be limited to maybe five or six usually surface level. Or if it is style, then I'd put all of these, I'd band these errors into one very generic error type. So that could be my issue really.

P5: But so you mentioned lack of experience in using the frame obviously there. So are you saying that if you were to do this again, it would be more straightforward to be easier,

P5: Yeah. Perhaps perhaps in terms of familiarity.

Interviewer: Okay. Okay, so you would've done eight Well, yeah. Overall, how easy or difficult was it for you to use the framework? And you said again, quite easy. Could you perhaps explain why you've chosen quite easy?

P5: I'd say quite easy rather than very easy because there was a lot of information for the marker to review before identifying the errors. And because it's a new code that I'm not, I'm not familiar with. It did take some time. And I think another reason I suppose why I put quite easy is not only the requirement to identify the category,

but also the column to a that two axes element was a bit of a challenge. I suppose the format as well, in terms of, yeah, just the amount of information, I suppose the stages that you go through. I'm just thinking back to doing the exercise. You read the sentence, you refer to the sections, you know, it takes a bit of time, and then there's a bit of umming and arring as to which one it fits in. Then you find, okay, let's say there's a spelling error there, so I've got it's spelling. I'll look for that. So it's an A3. And then not only that, I then have to look as to whether there's an omission. So it's quite time consuming in that respect. And that's just for one spelling error rather than the good old fashioned SP.

Interviewer: Yes, of course. Yeah. Yeah, I think let's talk about that. You any comments on how these might be used? And you said it's a little bit time consuming and it is of course more time consuming. Could you see a way that this could be used? I mean, maybe not every day

P5: Yeah, Yeah, I mean, I suppose it could be used. I'm thinking two groups from a teaching point of view, it's a good way to analyse errors in much more depth so you can find more about your learners, I would say. Because it's asking the teacher to not just put that SP for spelling, it's getting the teacher to look for what is the actual spelling error. Is it over inclusion or misselection? Then obviously you can identify patterns in that and identify that. Then on a broader scale, for teachers that could then link into your cohort of learners, is it an L1 issue? Or all these students making the same over use of spelling of the M letter. And that can definitely help then inform the teaching, because then, those specific errors can be dealt with in future classes. I think from the teacher's point of view, that's a benefit of the results and I guess if you pass it over to the students, the results could be used in a beneficial way for students. Because they would appreciate, I imagine if they got learner training on this error analysis. They would need some training, definitely. I think they would greatly benefit and appreciate it because again, it's more attention to their errors.

Interviewer: Thank you. Yeah. Do you have any other comments about, first of all, the depth of analysis? So let's say we've got the error codes in here. Filled them in. Is this a satisfactory depth of analysis state compared to what you would do normally with your error codes? Is it too much? Is it too little?

P5: If I were to take this to the language centre and share it with colleagues, I think there would be more people saying it's too much than too little. Because I think it's heavier than, than what I'm accustomed to from my language teaching where teachers' rooms, conversations and workshops on error correction. This is definitely more in depth than usual in terms of, yeah, so that's what I'd say in terms of comments, it's more than I would expect for an error analysis, but is that a bad thing? I don't think so. I think it's a good thing. But could it be, is there some kind of staging there that needs to go just slightly before this maybe? Is there, it is there more of an introductory stage before it that explains what errors are and what, why students make errors. And then going to the of slightly more staged approach. Perhaps rather than jumping into the heavy, the heavy meta language from straight in. Especially if you're talking about every day practising teachers who are

doing 20 hours of contact time per week. And to give them this on top of it...I'll stick with my SP.

Interviewer: Okay, well that partially answers comments on the guidance or framework or suggestions for improving the procedures. Number 14. a bit of a staged introduction to it. Yeah. You think of any other comments or ways of improving this? the processes, procedures, or the framework and guidance itself.

P5: Well, I think when I did this activity, it did make me real. It made me think about correcting learners work. And it made me think how I'm guilty of not doing it as in depth as I could have, as I should do, really. Because maybe it comes to the, the reality of workload and teachers. But it's easy if you're going to mark written work, to just skim it over and do a very simple correction code. And then what value does that have for the learner if they're just being told there's a generic error here in there, word order, let's say. I think something like this would definitely benefit TESOL practitioners as well as the learners. I guess it's finding a happy medium between what a lot of teachers doing now at the moment, which is a very easy level versus a far more in depth. But how deep do you go and how, how do you state that is a big question, which I guess I'm hoping you'll answer at the end of this research. Whether let's say an induction phase or an orientation phase of a programme. So I have a really busy summer programme where I have a week with teachers. And we talk about teaching approaches, methods, feedback, da, da, da. there's not one session on error analysis or correction. I have the time to deliver one. Something like that at the start of ELT programme would be really beneficial because we could do the teacher training and then we could do the learner training as well. But we don't have something clear and structured in place at the moment, so.

Interviewer: Right, right. Okay, great. Thank you very much. Is there anything else you want to say about it. Thank you very much. Participant five.

Appendix 7.1 Participant information sheet



**Manchester
Metropolitan**
University

Study Title Lexical Error Analysis (looking at the vocabulary errors in students' work). Participant Information Sheet 22/6/18 Version 1

I would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. Ask questions if anything you read is not clear or would like more information. Take time to decide whether or not to take part.

What is the purpose of the study? I'm researching the number and type of vocabulary errors in students' writing. This is part of my PhD.

Why have I been invited? I have devised a new framework for analysing lexical errors in students writing and I would like to test it on English Language Teachers.

Do I have to take part? It is up to you to decide. I will describe the study and then ask you to sign a consent form to show you agreed to take part. I will not use your name in the research and you will not be identifiable. You are under no obligation to participate, and I will not mention the names of the participants to anyone. You are free to withdraw at any time, without giving a reason. This will not affect you in any way.

What will happen to me if I take part? Nothing, but I will be very grateful.

Expenses and payments? There is no payment with this project.

What will I have to do? You only need to use the framework to analyse a short piece of writing and categorise some lexical errors. Then sign a consent form to show that you have agreed to do this.

What are the possible disadvantages and risks of taking part? There are none.

What are the possible benefits of taking part? You may learn, through experience, about how to analyse lexical errors systematically. The information I take from the study will help teachers to understand the types of errors that students make in their writing and may help to improve teaching and learning.

What if there is a problem? If you have a concern about any aspect of this study, you should ask to speak to the researcher who will do his best to answer your questions. You can contact me on a.picot@mmu.ac.uk or phone me on 0161 247 6183 if you have a problem or complaint with the research study. Alternatively, you could contact my Head of Department, Dr Derek Bousfield on d.bousfield@mmu.ac.uk or phone him on 0161 247 3620. Alternatively, if you wish to complain, you can find the University Complaints Procedure, here:

<https://www2.mmu.ac.uk/student-case-management/guidance-for-students/student-complaints-procedure/> My supervisor's name is Dr Huw Bell. He can be contacted on h.bell@mmu.ac.uk or 0161 247 6184. If you do not wish to discuss your complaint with the researcher, you should contact the supervisor in the first instance and then the College Research and Innovation (R&I) Manager. There are no compensation/insurance/indemnity schemes in place in the event of a complaint.

Will my taking part in the study be kept confidential? Your confidentiality will be safeguarded during and after the study by keeping your data on a password protected drive at all times. After the work has been completed, your name will be removed. These procedures match the principles in the Data Protection Act 1998. Electronic data will be stored on a password protected computer known only by researcher. The data will only be seen and used by myself and only used for my PhD. No further RGEC approval will be sought. The data will be kept for 2 years and then disposed of securely. All information which is collected about you during the course of the research will be kept strictly confidential, and any information about you which leaves the university will have your name removed so that you cannot be recognised.

What will happen if I don't carry on with the study? If you withdraw from the study we will destroy all your identifiable documents, but we will need to use the data collected up to your withdrawal.

What will happen to the results of the research study? I hope to publish them with my PhD. Results will be available to you on request. You will not be identified in any report/publication.

Who is organising or sponsoring the research? Funded by Manchester Metropolitan University.

Further information and contact details: If you have any further questions, please do not hesitate to contact me. A.picot@mmu.ac.uk. Many thanks for your participation!

**Appendix 7.2 Lexical error analysis of international students' compositions
consent form Version 2 (13/11/2018)**

Please initial all
boxes

I confirm that I have read and understand the information sheet dated
22/6/18 (version 1) for the above study.

I have had the opportunity to consider the information, ask questions and
have had these answered satisfactorily.

I understand that my participation is voluntary and that I am free to
withdraw at any time without giving any reason.

I agree to take part in the above study.

Participant number: _____

Name of Participant:

Date:

Name of Person taking consent. (leave blank)

Date (leave blank)

Appendix 7.3 Printed or emailed instructions for data collection Participant number: _____

Task 1) Identification of Lexical Errors. Underline all the lexical errors in the essay. (A written lexical error is a lexical form or combination of forms which would not be made by a highly-skilled English user, in the same genre). Follow the rules below to **exclude grammatical errors from the analysis**:

Grammar Error	Examples	Advice
Article errors	<i>a/an/the/article omission</i>	In fixed phrases, such as 'What's <i>*a matter?</i> ' Here, the error is collocational and is therefore a lexical error.
Inflectional affixes	<i>work/worked/working or dog/dogs</i>	Do include derivational affix errors (e.g. <i>He is very consider*able</i>)
Infinitive 'to' errors	<i>*Live without love is not *live</i>	
Gerund errors	<i>I like ski*/I like to ski*ing</i>	
Genitive errors (including possession errors)	<i>It's Tom* bike. It is *the bike of Tom.</i>	
Comparative and superlative errors	<i>It is *expensiver, It is the *most big</i>	
Clause errors	<i>It's not difficult *for getting to a hospital and *While waiting, my hamburger went cold</i>	
Relative clause errors (including incorrect relative pronoun and referential errors).	<i>The man *what lives next door is a doctor.</i>	
Countability Errors	<i>Ten items or *less. There are too *much cars.</i>	
Tense errors	<i>Yesterday I *go to school.</i>	

Ignore errors with punctuation, including capitalisation. (**vienna is *In Austria*). All other errors are lexical.

Separating Lexical Errors Multiple lexical errors in a phrase are counted separately. However, when a sentence lacks coherence, and it is impossible to separate errors, count them as one per phrase. E.g. '**all people there aren't in the globe next a few years*', there are three, based on the three phrases (noun phrase, verb phrase and prepositional phrase).

Please underline all lexical errors in the essay below.

Essay question: To what extent do you agree that reading is the most important skill in university study?

To get a perfect academic performance is a basic quality that every student wants. So, students must develop various skills to achieve the goal. There are many skills to help students to study. Some people think the Reading skill is the most important compared with others. However, others hold the opposite view. Let me discuss the topic and indicate how I disagree with the topic.

Reading skill, undoubtedly, is a very crucial factor to master for every student. As a student who finishes a lot of tasks and homework, I must use a lot of knowledge and documents that I do not know to make it better. In this case students will read different kinds of books, essays, files and so on. Thus, how to read them fast and correctly is a fundamental ability to finish tasks. Though reading skill has many advantages. Is it the most important? I do not think so.

Not only reading skill can help students do better but also other academic skills will help more. Such as writing, listening, speaking and so on. Firstly, writing skill is a way to show students' views. As everyone knows, a paper or an essay must be finished in the end of a semester. So, what score can a student obtain is determined by how to organise words to indicate points. So, in my opinion, it is the most direct approach to gain great academic performance. Then, the listening skill is a prior skill for students. Before you use the new knowledge, you need to hear the content of what the professor says clearly. No tutor teaches students only by writing on the blackboard without any voice. So, listening is a very effective way to obtain new things. As for speaking, it is an imperative skill for every foreign student to own. Speaking skill is a communicative skill. Who wants to be a silent person? Whatever a student hears and writes and how fast a student reads, if a student cannot talk about their own views with others, it is vain for the student to have so many good skills. Speaking to others and getting more information from others is a correct way to study in an academy and improve one's ability.

Above all, I think, the most important is not only reading skill. Other skills are not either. The integrated skill is the best way to obtain the biggest progress and a better academic performance.

THANK YOU FOR PARTICIPATING!

Participant number: _____

Task 2) Lexical Error Identification Task

- Look at the errors (in italics) below and decide which type of lexical errors they are according to the codes in the corresponding framework cells below. Record the codes in the P Code column.
- Allocate a confidence score (one per error) in the P Conf column:
 - 3 = **Completely sure** that this is the only way to categorise this error.
 - 2 = **Not completely sure** that this is the only way to categorise this error.
 - 1 = **Unsure**. There is more than one way to categorise this error.

Lexical Errors (in italics) and Participant's Categorisations and Confidence Scores

No		P_Code	P_Conf
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.		
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.		
3	perfect academic performan is a <i>basic quality</i> that every student wants.		
4	<i>Let me discuss the topic</i>		
5	and indicate how I disagree with the <i>topic</i> .		
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master		
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master		
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master		
9	As a student who <i>finishes</i> amount of tasks and homework,		
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .		

11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .		
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .		
13	different kinds of books <i>essaies</i> , files and so on.		
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .		
15	an essay must be finished <i>in</i> the end of a semester.		
16	<i>Is it the most important ?</i> I do not think so.		
17	Is it the most important ? <i>I do not think so</i> .		
18	So, what score can a student obtain is determined by <i>how to organise</i> words to indicate points		
19	So, <i>in my opinion</i>		
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .		
21	<i>to gain great academic performance</i>		
22	, the listening skill is <i>a prior skill</i> for a students.		
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.		
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.		
25	content of what proffesser says clearly.		
26	content of what proffesser says clearly.		
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .		

28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .		
29	Who <i>wannna</i> be a silent person ?		
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills		
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills		
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills		
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>		
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.		
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .		
36	I think, the most important is not only reading skill. <i>Other skills are not. either</i> .		
37	the best way to <i>obtain the biggest</i> progress		

Confidence scores: 3=Very confident about accuracy of allocation. 2= Quite confident. 1=Not Confident.

Lexical Error Analysis Framework V2

Omission = omitted items that are required
 Over-inclusion = extra items that should not be there
 Mis-selection = the wrong items have been chosen
 Mis-order = items are in the wrong order
 Blend= two **correct (not incorrect)** items have been incorrectly combined in some way

Error type	Description/advice	Examples	O-MISSION	OVERIN-CLUSION	MIS-SELEC-TION	MIS-ORDER	BLEND
Section A Form							
A1 SUFFIX	Correct root, but incorrect suffix. Do not include verb tense errors (He was go) subject/verb agreement errors (He go) or errors with plurals	<i>They were very consider*able</i> <considerate> <i>The data were analys*t.</i> <analysed>	1	2	3	N/A	4
A2 PREFIX	Correct root, but incorrect/missing prefix.	<i>It's quite *inimportant.</i> <unimportant>	5	6	7	N/A	8
A3 SPELLING	Meaning of intended word is clear and exists in English. Words with unclear meaning should be categorised as B1 .	<i>I need to *safe some money.</i> <save> <i>Put the *folwers in water.</i> <flowers>	9	10	11	12	13
A4 TWO WORDS	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-law or	<i>*Every one will pay.</i> <everyone>	N/A	N/A	N/A	N/A	14

SHOULD BE ONE	ice cream/ ice-cream). Does not include an extra incorrect word (see Section B2).	<i>I found it *on line.</i> <online>							
A5 ONE WORD SHOULD BE TWO.	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-law or ice cream/ ice-cream). Does not include missing word (see Section B2).	<i>Where is the *paperbin?</i> <paper bin> <i>Put it on the *dinnertable.</i> <dinner table>	N/A	N/A	N/A	N/A	15		
Section B Meaning			O-MISSION	OVERIN-CLUSION	MISSELEC-TION	MIS-ORDER	BLEND		
B1 Co-HERENCE	Words do not exist in English or meaning of whole phrase is unclear. Only select this category if you have re-read and still cannot be sure of the meaning of the word or phrase.	<i>*Where is those that are? <?></i> <i>I am *lipsh <?></i>	16						
Errors with Single Words (if clearly a misspelling, allocate to A3, e.g. <i>I need to *safe some money.</i> <save>)									
B2 SINGLE WORD (MISC)	An error with a single word (error could be related or unrelated to correct word). Word exists in English. Does not fit into other B2 categories below.	Exclaimer???	17	18	19	N/A	20		
B2A CON-JUNCTION	Include all conjunctions or transition signals here, including multiword items, such as 'in spite of this'.	<i>It was hot *and I took off my coat.</i> <so> <i>In *despite of the rain, we went swimming.</i> <spite>	21	22	23	N/A	24		

B2B NOUN	Include compound nouns, but not multi-word units/collocations/fixed expressions.	<i>It was hot so I took off my *hand. <coat?></i> <i>We visited the *art museum. <art gallery></i>	25	26	27	N/A	28
B2C ADJECTIVE		<i>I feel *serious. <stressed></i> <i>The views were *handsome. <beautiful></i>	29	30	31	N/A	32
B2D ADVERB		<i>He sings *deliciously.</i> <i>They are growing *quick. <quickly></i>	N/A	33	34	N/A	35
B2E PRE-POSITION	Prepositions of time and place and dependent propositions, not errors with infinitive 'to'. Include incorrect phrasal verb errors with particles/prepositions here.	<i>I woke up *on 6am. <at></i> <i>See you *in the corner. <at/on></i> <i>I was suspicious *about his behaviour. <of></i>	36	37	38	N/A	39
B2F PRONOUN		<i>My mother, *she is ill.</i>	40	41	42	43	N/A
B2G VERB	Include incorrect phrasal verb errors with roots verbs here.	<i>Do you *get the time, please? <have></i> <i>He *put on the table. <laid></i>	44	45	46	47	48

Errors that span across more than one word			O-MISSION	OVERINCLUSION	MISSELECTION	MIS-ORDER	BLEND
B3 PHRASE ERROR (MISC)	Meaning is clear, but word choices are non-expert-like. There may be more than one error in the phrase (or multi-word unit). Also use for whole problematic phrases which require re-writing.	* <i>The number of people is really great.</i> <There is a large number of people>. * <i>Next a few years.</i> <In the next few years.> <i>Oranges taste *not nice, <awful></i>	N/A	N/A	49	50	51
B3A Mis-ORDERING IN FIXED PHRASES	Collocational errors relating to order of pairs of words or groups of three.	* <i>hike-hitch</i> <hitch-hike> * <i>fro and to</i> <to and fro> * <i>Dark, tall and handsome</i> <Tall, dark and handsome>	52				
B3B VERBOSITY	Too many words are used, or points are repeated unnecessarily. Categorise single extra word in B2 overinclusion above.	<i>He bought an apple and* he bought a banana.</i> <He bought an apple and a banana>. <i>I like cake. *I like cake because it is sweet.</i> <I like cake because it is sweet>	53				
B3C UNDER	Phrases/sentences that require more detail to improve expression. Categorise single missing word	<i>Although *cars in the country are lower...</i> <Although there	54				

SPE- CIFICATION	in B2 omission above. If meaning is unclear, categorise as B1.	are lower car numbers in the country....>). <i>We cut *because tree high.</i> <We cut the trees because they were getting too high.>	
B3D FORMALITY	Phrases that are correct but seem overly formal or informal for the genre. If there is a formality error with a single word, allocate to word class section B2 above.	<i>I *informed my girlfriend via the medium of the telephone.</i> <I told my girlfriend on the phone.> <i>*Milk stocks in the refrigerator are starting to ebb.</i> <We need more milk.>	55
B4 UNCATEGO- RISABLE	Only use this category if the error cannot fit into any category above.	?	56

- If the meaning of the error is unclear, consult the full original composition.
- Categorise the error based on what the student wrote, not what they should have written. E.g. If a student writes *Oranges taste *not nice*, meaning, Oranges taste awful, this would be a B3 phrase misselection error, code 49, not a B2D adverb error. The student made an error with these two words, not the missing adverb, so it should be classified as such. Categorise the written error. Not what they should have written.
- If the meaning of the error still cannot be understood, allocate it to the coherence category.
- Avoid speculation of what caused the error. Simply focus on the actual error. You may consider causality later.
- If you cannot categorise the error, use the final category: 'Uncategorisable'.

Appendix 7.4 Letter of ethical approval



15/01/2019

Project Title: Lexical Error Analysis of International Students' Compositions

EthOS Reference Number: 0906

Ethical Opinion

Dear Anthony Picot,

The above application was reviewed by the Arts and Humanities Research Ethics and Governance Committee and, on the 15/01/2019, was given a favourable ethical opinion. The approval is in place until 31/07/2020 .

Conditions of favourable ethical opinion

Application Documents

Document Type	File Name	Date	Version
Additional Documentation	MMU-Application-for-Ethical-Approval-September-2015	03/06/2018	1
Additional Documentation	MMU-Research-Insurance-Checklist-v1-0-19-Sept-2016	03/07/2018	1
Additional Documentation	risk assessment_07032018_135302	03/07/2018	1
Additional Documentation	Interview questions for staff	03/07/2018	1
Additional Documentation	Interview Questions for students	03/07/2018	1
Project Proposal	MMU-Application-for-Ethical-Approval-September-2015	18/07/2018	1
Consent Form	Consent form for students	18/07/2018	1
Information Sheet	Partipant-Information-Sheet for staff	18/07/2018	1
Consent Form	Consent form for Staff	04/10/2018	2
Additional Documentation	Consent form for Staff	13/11/2018	2
Additional Documentation	Consent form for students	27/12/2018	3
Additional Documentation	Partipant-Information-Sheet for staff	27/12/2018	3
Additional Documentation	Partipant-Information-Sheet for students	27/12/2018	3

The Arts and Humanities Research Ethics and Governance Committee favourable ethical opinion is granted with the following conditions

Adherence to Manchester Metropolitan University's Policies and procedures

This ethical approval is conditional on adherence to Manchester Metropolitan University's Policies, Procedures, guidance and Standard Operating procedures. These can be found on the Manchester Metropolitan University Research Ethics and Governance webpages.

Amendments

If you wish to make a change to this approved application, you will be required to submit an amendment. Please visit the Manchester Metropolitan University Research Ethics and Governance webpages or contact your Faculty research officer for advice around how to do this.

We wish you every success with your project.

Art and Humanities Research Ethics and Governance Committee

Appendix 7.5 Questionnaire

1) What is your nationality?

2) What is your first language?

3) What is your status? (Tick all that apply)

BATESOL Student

MATESOL Student

Experienced English Language Teacher (with minimum CELTA or Trinity Cert TESOL)

University TESOL/Linguistics Lecturer

4) Have you studied an aspect of TESOL/Linguistics? If so, which? (pls underline all that apply)

None / CELTA/Cert TESOL / Online TESOL qualification / BA / MA /Diploma/ PhD

Other (pls specify) _____

5) Are you currently studying an aspect of TESOL/Linguistics? If so, pls underline all that apply

None / CELTA/Cert TESOL / Online TESOL qualification / BA / MA /Diploma/ PhD

Other (pls specify) _____

6) If you are an English Language Teacher, or have worked as one, how many years' experience do you have (pls underline your response)?

0-5 6-10 11-15 16-20 21-25 Over 25 years

Appendix 7.6 Emails to participants

Email 1

Hi X

Thanks so much for agreeing to help me out with my data collection. I'm so very grateful. This shouldn't take more than 15 mins. I'd be grateful if you could Identify all lexical errors in a piece of student writing.

Please could you complete the tasks in the following order.

1. Read the Participant Information Sheet (attachment 1). If you are happy to participate, save the document to your computer and complete the consent form and questionnaire, which are in the same document. (make sure the docs are editable and not in read only mode)
2. Click on this link to listen to the instructions for the first task
https://mmutube.mmu.ac.uk/media/Lexical+Error+Analysis+Identification+of+Lexical++Errors/1_yvzqt53q
3. Save the Lexical Error Identification Task doc to your computer and complete the first task (make sure the docs are editable and not in read only mode).
4. Return the completed Questionnaire, Consent form and Task sheet 1 to mw via email.

Many thanks again!

Tony

Appendix 7.7 Email 2

Thanks so much for agreeing to help me out with my data collection. I'm so very grateful. This shouldn't take more than 30 mins. I'd be grateful if you could categorise a list of lexical errors using a new framework.

1. Click on this link to listen to the instructions for the second task.
https://mmutube.mmu.ac.uk/media/Lexical+Error+Analysis+Error+Categorisation/1_amt0uey1
2. Save the Lexical Error Categorisation Task doc to your computer and complete the second task. (add in the doc the request not to make changes to responses to task 1 as error identification can be subjective and I am interested in the differences) (make sure the docs are editable and not in read only mode)
3. Ensure that you have saved changes to all three docs that you have downloaded and email them back to me. I will ensure that anonymity is maintained.

Many thanks again!

Tony

Appendix 7.8 Table 7.3 Part One Error Identification

N o	Error	R	P1/P2	P3/P4	P5/P6	Agr %
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	1	1	1	1	100
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	1	1	1		75
3	perfect academic performan is a <i>basic quality</i> that every student wants.	1	1	1		75
4	<i>Let me discuss the topic</i>	1			1	50
5	and indicate how I disagree with the <i>topic</i> .	1		1		50
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	1		1		50
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	1	1			50
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	1		1	1	75
9	As a student who <i>finishes</i> amount of tasks and homework,	1	1	1	1	100
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	1			1	50
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .	1	1		1	75
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .	1	1			50
13	different kinds of books <i>essaies</i> , files and so on.	1	1			50
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .	1			1	50
15	an essay must be finished <i>in</i> the end of a semester.	1	1	1	1	100
16	<i>Is it the most important ?</i> I do not think so.	1			1	50
17	Is it the most important ? <i>I do not think so</i> .	1			1	50
18	So, what score can a student obtain is determined by <i>how to organise words</i> to indicate points	1				25
19	So, <i>in my opinion</i>	1				25
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	1				25
21	to <i>gain great academic performance</i>	1	1		1	75

22	, the listening skill is a <i>prior skill</i> for a students.	1	1	1		75
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	1			1	50
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	1				25
25	content of what proffesser says clearly.	1	1			50
26	content of what proffesser says clearly.	1	1			50
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	1	1	1	1	100
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	1	1	1	1	100
29	Who <i>wannna</i> be a silent person ?	1	1		1	75
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not...	1				25
31	...if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	1	1			50
32	...if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	1	1	1		75
33	...if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	1	1	1	1	100
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	1	1	1		75
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .	1				25
36	I think, the most important is not only reading skill. <i>Other skills are not. either</i> .	1				25
37	the best way to <i>obtain the biggest</i> progress	1	1	1	1	100
38	compared <u>with</u> others.		1	1		50
39	<u>files</u> and so on		1	1		50
40	how to read them <u>fast</u>		1	1		50
41	is a <u>fundamental ability</u> to finish		1	1		50
42	There are many <u>skills</u> to help students			1		25

43	Some people think <u>the Reading skill</u>			1		25
44	is a <u>correct way</u> to study			1	1	50
45	<u>Then</u> , the listening skill is a			1		25
46	Firstly, <u>writing skill</u> is a way to show students' views			1		25
47	Not only reading skill can <u>help</u> student doing				1	25
48	<u>As everyone know</u> , a paper or an				1	25
49	No tutor teaches students <u>only</u> by writing				1	25
50	for every <u>foreign</u> student				1	25
51	<u>As for speaking</u> , it is a imperative skill				1	25
52	be a <u>silent person</u> ?				1	25
53	Speaking to others and <u>get</u> more information from others				1	25
54	Above all, <u>I think</u> , the most important				1	25
55	and a <u>better</u> academic performance.				1	25
	Totals	37	25	24	27	51.36

Appendix 7.9 Table 7.4 Part One error categorisation and confidence scores

Error No	R code	R Conf	P1 Code	P1 Conf	P2 Code	P2 Conf	P3 Code	P3 Conf	P4 Code	P4 Conf	P5 Code	P5 Conf	P6 Code	P6 Conf	% agr	ave conf
1	46	3	46	3	46	3	46	3	46	3	46	3	46	3	100	3.00
2	9	3	9	3	9	3	9	2	3	2	1	3	1	3	57	2.71
3	49	3	49	3	49	3	49	3	49	3	49	2	49	2	100	2.71
4	55	3	55	3	55	3	55	2	55	2	55	3	55	3	100	2.71
5	27	3	19	1	19	2	27	2	27	3	27	3	27	3	71	2.43
6	26	3	26	1	26	1	26	3	26	3	14	3	14	3	71	2.43
7	9	3	9	3	9	3	9	3	9	3	9	3	9	3	100	3.00
8	27	3	27	3	27	3	27	3	27	3	27	3	27	3	100	3.00
9	46	3	49	2	49	2	19	3	19	3	46	3	46	3	43	2.71
10	55	3	55	3	55	3	55	3	55	3	55	3	55	3	100	3.00
11	54	3	54	2	54	2	54	2	54	3	53	3	53	3	71	2.57
12	54	3	54	2	54	3	54	3	54	3	49	3	49	3	71	2.86
13	11	3	11	3	11	3	11	3	11	3	3	3	3	3	71	3.00
14	55	3	55	3	55	3	55	3	55	3	53	1	53	1	71	2.43
15	38	3	38	3	38	3	38	3	38	3	38	3	38	3	100	3.00
16	55	3	55	3	55	3	54	3	54	3	55	3	55	3	71	3.00
17	55	3	55	3	55	3	55	3	55	3	55	3	55	3	100	3.00
18	49	3	49	2	49	3	49	3	49	3	53	2	53	2	71	2.57
19	55	3	55	3	55	3	55	3	55	3	55	3	55	3	100	3.00
20	49	3	49	2	49	2	49	3	49	2	54	1	54	1	71	2.00
21	49	3	49	2	49	3	49	3	49	3	49	3	49	3	100	2.86
22	49	3	49	2	49	3	54	2	54	2	49	3	49	3	71	2.57
23	55	3	55	3	55	3	55	3	55	3	55	3	55	3	100	3.00

24	46	3	46	3	46	3	46	3	46	3	46	3	46	3	100	3.00
25	10	3	10	3	10	3	10	3	10	3	13	3	13	3	71	3.00
26	11	3	11	3	11	3	11	3	11	3	11	3	11	3	100	3.00
27	49	3	49	2	49	3	54	2	54	2	49	3	49	3	71	2.57
28	46	3	46	3	46	3	46	3	46	3	46	3	46	3	100	3.00
29	55	3	20	2	20	2	55	3	55	3	55	3	55	3	71	2.71
30	23	2	23	2	23	2	19	2	19	2	55	3	55	3	43	2.29
31	14	3	14	3	14	3	14	3	14	3	14	3	14	3	100	3.00
32	46	3	36	2	36	3	46	3	46	3	46	3	46	3	71	2.86
33	49	3	49	3	49	3	49	3	49	3	16	3	16	3	71	3.00
34	54	2	54	3	54	3	49	2	16	2	1	3	1	3	43	2.57
35	16	3	50	3	50	3	16	2	16	3	16	3	16	3	71	2.86
36	16	3	16	2	16	3	16	3	16	3	50	1	50	1	71	2.29
37	49	3	49	3	49	3	49	3	49	3	55	2	55	2	71	2.71
Averages		2.95		2.57		2.78		2.76		2.81		2.76		2.76	80.11	2.77

No	Error	R Code	P1 Code	P2 Code	P3 Code	P4 Code	P5 Code	P6 Code	Mode
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	49	49	49	49	49	54	54	49
21	to <i>gain great academic performance</i>	49	49	49	49	49	49	49	49
22	, the listening skill is <i>a prior skill</i> for a students.	49	49	49	54	54	49	49	49
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	55	55	55	55	55	55	55	55
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	46	46	46	46	46	46	46	46
25	content of what proffesser says clearly.	10	10	10	10	10	13	13	10
26	content of what proffesser says clearly.	11	11	11	11	11	11	11	11
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	49	49	49	54	54	49	49	49
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	46	46	46	46	46	46	46	46
29	Who <i>wannna</i> be a silent person ?	55	20	20	55	55	55	55	55
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills	23	23	23	19	19	55	55	23
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	14	14	14	14	14	14	14	14
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	46	36	36	46	46	46	46	46
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	49	49	49	49	49	16	16	49
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	54	54	54	49	16	1	1	54
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .	16	50	50	16	16	16	16	16
36	I think, the most important is not only reading skill. <i>Other skills are not. either</i> .	16	16	16	16	16	50	50	16
37	the best way to <i>obtain the biggest</i> progress	49	49	49	49	49	55	55	49

Appendix 7.11 Table 7.6 Part Two Group A and Group B error identification

E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB	E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB	E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB	E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB
1	10	48	4	24	24	6	29	3	18	47	1	5	1	6	70	3	14	2	12
2	20	95	13	76	25	15	71	10	59	48	7	33	3	18	71	1	5	3	18
3	9	43	6	35	26	15	71	11	65	49	2	10	1	6	72	0	0	1	6
4	8	38	3	18	27	14	67	10	59	50	4	19	4	24	73	3	14	2	12
5	9	43	3	18	28	16	76	10	59	51	1	5	1	6	74	4	19	1	6
6	4	19	4	24	29	17	81	13	76	52	5	24	3	18	75	1	5	1	6
7	14	67	9	53	30	4	19	5	29	53	4	19	2	12	76	4	19	4	24
8	9	43	8	47	31	4	19	5	29	54	3	14	2	12	77	7	33	3	18
9	11	52	4	24	32	16	76	10	59	55	2	10	1	6	78	4	19	2	12
10	6	29	7	41	33	20	95	13	76	56	1	5	6	35	79	3	14	1	6
11	11	52	8	47	34	19	90	16	94	57	0	0	1	6	80	2	10	1	6
12	11	52	6	35	35	8	38	5	29	58	2	10	5	29	81	2	10	0	0
13	16	76	13	76	36	12	57	5	29	59	0	0	2	12	82	4	19	0	0
14	5	24	4	24	37	14	67	9	53	60	1	5	2	12	83	4	19	0	0
15	16	76	6	35	38	5	24	4	24	61	8	38	10	59	84	2	10	1	6
16	4	19	4	24	39	4	19	2	12	62	5	24	3	18	85	3	14	0	0
17	4	19	1	6	40	7	33	1	6	63	10	48	7	41	86	1	5	2	12
18	7	33	6	35	41	4	19	6	35	64	5	24	4	24	87	0	0	0	0
19	2	10	1	6	42	1	5	2	12	65	4	19	1	6	88	0	0	0	0
20	5	24	1	6	43	1	5	2	12	66	6	29	9	53	89	1	5	0	0
21	13	62	8	47	44	9	43	4	24	67	3	14	2	12	90	6	29	5	29
22	16	76	9	53	45	6	29	1	6	68	1	5	0	0	91	2	10	1	6
23	4	19	4	24	46	3	14	2	12	69	6	29	1	6	92	1	5	0	0

E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB	E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB	E No	No GA agreed	% Agree GA	No G B agreed	% Agree GB
93	2	10	1	6	117	1	5	1	6	143	1	5	1	6
94	2	10	2	12	118	1	5	0	0	144	0	0	1	6
95	4	19	3	18	119	2	10	1	6	145	1	5	1	6
96	4	19	1	6	120	1	5	1	6	146	1	5	0	0
97	2	10	1	6	121	5	24	0	0	147	2	10	0	0
98	5	24	4	24	122	3	14	1	6	148	2	10	0	0
99	1	5	1	6	123	0	0	2	12	149	2	10	0	0
100	0	0	1	6	124	0	0	1	6	150	1	5	0	0
101	4	19	2	12	125	0	0	1	6	151	0	0	0	0
102	2	10	3	18	126	2	10	2	12	152	0	0	1	6
103	4	19	2	12	127	1	5	0	0	153	0	0	1	6
104	1	5	2	12	128	1	5	0	0	154	0	0	1	6
105	3	14	1	6	129	1	5	0	0					
106	2	10	1	6	130	2	10	1	6					
107	2	10	3	18	131	1	5	0	0					
108	2	10	1	6	132	1	5	0	0					
109	1	5	1	6	133	0	0	1	6					
110	2	10	0	0	134	0	0	1	6					
111	3	14	2	12	135	1	5	0	0					
112	2	10	0	0	136	1	5	0	0					
113	2	10	1	6	137	0	0	1	6					
114	2	10	0	0	138	0	0	2	12					
115	1	5	0	0	140	1	5	1	6					
116	1	5	1	6	141	1	5	0	0					
139	1	5	1	6	142	1	5	0	0					

Appendix 7.12 Table 7.7 Part Two Group A error identification

No	Error	GA No agreed	GA % Agr
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	10	47.6
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	20	95.2
3	perfect academic performan is a <i>basic quality</i> that every student wants.	9	42.9
4	<i>Let me discuss the topic</i>	8	38.1
5	and indicate how I disagree with the <i>topic</i> .	9	42.9
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	4	19
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	14	66.7
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	9	42.9
9	As a student who <i>finishes</i> amount of tasks and homework,	11	52.4
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	6	28.6
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .	11	52.4
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .	11	52.4
13	different kinds of books <i>essaies</i> , files and so on.	16	76.2
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .	5	23.8
15	an essay must be finished <i>in</i> the end of a semester.	16	76.2
16	<i>Is it the most important ?</i> I do not think so.	4	19
17	Is it the most important ? <i>I do not think so</i> .	4	19
18	So, what score can a student obtain is determined by <i>how to organise words</i> to indicate points	7	33.3
19	So, <i>in my opinion</i>	2	9.52
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	5	23.8
21	<i>to gain great academic performance</i>	13	61.9
22	, the listening skill is a <i>prior skill</i> for a students.	16	76.2
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	4	19
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	6	28.6
25	content of what proffesser says clearly.	15	71.4
26	content of what proffesser says clearly.	15	71.4
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	14	66.7
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	16	76.2
29	Who <i>wannna</i> be a silent person ?	17	81

30	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills	4	19
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	4	19
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	16	76.2
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	20	95.2
34	Speaking to others and get more information from others is a correct way to <i>study academy</i> and improve one's ability.	19	90.5
35	I think, <i>the most important is not only reading skill.</i> Other skills are not. either .	8	38.1
36	I think, the most important is not only reading skill. <i>Other skills are not. either .</i>	12	57.1
37	the best way to <i>obtain the biggest</i> progress	14	66.7
38	compared <u>with</u> others.	5	23.8
39	<u>files</u> and so on	4	19
40	how to read them <u>fast</u>	7	33.3
41	is a <u>fundamental ability</u> to finish	4	19
42	There are many <u>skills</u> to help students	1	4.76
43	Some people think <u>the Reading skill</u>	1	4.76
44	is a <u>correct</u> way to study	9	42.9
45	<u>Then</u> , the listening skill is a	6	28.6
46	Firstly, <u>writing skill</u> is a way to show students' views	3	14.3
47	Not only reading skill can <u>help</u> student doing	1	4.76
48	<u>As everyone know</u> , a paper or an	7	33.3
49	No tutor teaches students <u>only</u> by writing	2	9.52
50	for every <u>foreign</u> student	4	19
51	<u>As for</u> speaking, it is a imperative skill	1	4.76
52	be a <u>silent person</u> ?	5	23.8
53	Speaking to others and <u>get</u> more information from others	4	19
54	Above all, <u>I think</u> , the most important	3	14.3
55	and a <u>better</u> academic performance.	2	9.52
56	<u>Though</u> reading skill has many advantages	1	4.76
58	<u>I must</u> use lots	2	9.52
60	the content of what proffesser <u>says clearly</u> .	1	4.76
61	way to obtain new <u>things</u> .	8	38.1
62	and <u>indicate how</u> I disagree	5	23.8
63	who finishes <u>amount</u> of tasks and homework,	10	47.6

64	<u>Not only reading skill can help student doing</u>	5	23.8
65	So, what score <u>can a student obtain</u>	4	19
66	how to organise words to <u>indicate points</u>	6	28.6
67	to have so many <u>good skills</u>	3	14.3
68	Speaking to others <u>and get more information</u>	1	4.76
69	<u>the integrated skill is the best way</u>	6	28.6
70	, the listening skill is a prior skill for <u>a students.</u>	3	14.3
71	As for speaking, it is <u>a imperative skill for every foreign student to own.</u>	1	4.76
73	read them fast and <u>correctly</u>	3	14.3
74	Not only reading skill can help student <u>doing better</u>	4	19
75	but <u>also other academic skills</u>	1	4.76
76	but also other academic skills will <u>help more</u>	4	19
77	it is a <u>imperative skill for every foreign student to own.</u>	7	33.3
78	students must develop various skills to achieve <u>the goal</u>	4	19
79	<u>There are many skills to help students to study</u>	3	14.3
80	and indicate how I <u>disagree with the topic.</u>	2	9.52
81	Not only reading skill can help student <u>doing</u>	2	9.52
82	<u>Such as writing, listening</u>	4	19
83	Whatever a student <u>hear and write</u>	4	19
84	Whatever a student <u>hear and write</u>	2	9.52
85	and how fast a student <u>read</u>	3	14.3
86	So, <u>what</u> score can a student obtain	1	4.76
89	There are many skills <u>to help students to study</u>	1	4.76
90	is a <u>very crucial</u>	6	28.6
91	Whatever a student hear and write and <u>how fast a student read, if a student</u>	2	9.52
92	get more information from others is <u>a correct way</u>	1	4.76
93	Thus, <u>how</u> to read them fast and	2	9.52
94	<u>No tutor teaches students</u>	2	9.52
95	very effective way to <u>obtain new things</u>	4	19
96	<u>Above all, I think, the most important</u>	4	19
97	So, what <u>score</u> can a student obtain	2	9.52
98	fundamental ability <u>to finish tasks.</u>	5	23.8
99	that every <u>student wants.</u>	1	4.76
101	Though reading <u>skill</u> has many advantages.	4	19
102	, writing skill is a way to <u>show students' views</u>	2	9.52
103	<u>So, in my opinion, it is the most</u>	4	19
104	listening is a <u>very effective way</u>	1	4.76
105	<u>Speaking skill is communicative skill.</u>	3	14.3
106	skill is the <u>best way</u> to obtain	2	9.52
107	I must <u>use</u> lots of knowledge and	2	9.52
108	I must use lots of <u>knowledge</u> and	2	9.52
109	how to organise words to indicate <u>points</u>	1	4.76
110	develop various skills to achieve <u>the goal.</u>	2	9.52
111	, <u>a paper or an essay must be finished</u>	3	14.3

112	<u>the listening skill</u>	2	9.52
113	the listening <u>skill</u>	2	9.52
114	Before you <u>use the new knowledge,</u>	2	9.52
115	As for speaking, it is a imperative <u>skill</u>	1	4.76
116	<u>and a better academic performance.</u>	1	4.76
117	think the <u>Reading skill</u> is the most important	1	4.76
118	There are many skills <u>to help students</u>	1	4.76
119	There are many skills to help students <u>to study</u>	2	9.52
120	As a student who finishes amount of <u>tasks</u>	1	4.76
121	<u>In this case</u> students will read different kinds of books	5	23.8
122	Though reading skill <u>has many advantages.</u>	3	14.3
126	<u>Such as writing, listening, speaking and so on.</u>	2	9.52
127	is the most important compared with <u>others.</u>	1	4.76
128	listening is a very <u>effective way</u>	1	4.76
129	<u>As for speaking,</u> it is a imperative skill	1	4.76
130	to have <u>so many</u> good skills	2	9.52
131	others and get <u>more</u> information from others	1	4.76
132	the integrated skill is the <u>best</u> way	1	4.76
135	compared with <u>others.</u>	1	4.76
136	<u>get more information from others is</u>	1	4.76
139	is a way to show students' <u>views.</u>	1	4.76
140	<u>to master for every student.</u>	1	4.76
141	<u>Thus,</u> how to read them fast and	1	4.76
142	<u>Firstly,</u> writing skill is a way to show students' views	1	4.76
143	<u>Speaking to others and get more information from others</u>	1	4.76
145	<u>So,</u> students must develop various skills	1	4.76
146	indicate <u>how</u> I disagree with the	1	4.76
147	the <u>content</u> of what	2	9.52
148	and improve one's <u>ability.</u>	2	9.52
149	discuss <u>the</u> topic	2	9.52
150	<u>how to read them fast and correctly is a fundamental ability to finish tasks.</u>	1	4.76
	Totals		20.6 (ave)

Appendix 7.13 Table 7.8 Part Two Group B error identification

No	Error	No GB Agr	% GB Agr
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	4	23.53
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	15	88.24
3	perfect academic performan is a <i>basic quality</i> that every student wants.	5	29.41
4	<i>Let me discuss the topic</i>	2	11.76
5	and indicate how I disagree with the <i>topic</i> .	5	29.41
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	2	11.76
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	11	64.71
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	7	41.18
9	As a student who <i>finishes</i> amount of tasks and homework,	5	29.41
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	6	35.29
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .	9	52.94
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .	6	35.29
13	different kinds of books <i>essaies</i> , files and so on.	13	76.47
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .	2	11.76
15	an essay must be finished <i>in</i> the end of a semester.	8	47.06
16	<i>Is it the most important ?</i> I do not think so.	2	11.76
17	Is it the most important ? <i>I do not think so</i> .	1	5.88
18	So, what score can a student obtain is determined by <i>how to organise words</i> to indicate points	6	35.29
19	So, <i>in my opinion</i>	1	5.88
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance .	1	5.88
21	<i>to gain great academic performance</i>	9	52.94
22	, the listening skill is a <i>prior skill</i> for a students.	10	58.82
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	2	11.76
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	4	23.53
25	content of what proffesser says clearly.	11	64.71
26	content of what proffesser says clearly.	11	64.71
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	8	47.06
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	11	64.71

29	Who <i>wannna</i> be a silent person ?	14	82.35
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills	3	17.65
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	5	29.41
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	12	70.59
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	13	76.47
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	16	94.12
35	I think, <i>the most important is not only reading skill.</i> Other skills are not. either .	4	23.53
36	I think, the most important is not only reading skill. <i>Other skills are not. either .</i>	5	29.41
37	the best way to <i>obtain the biggest</i> progress	10	58.82
38	compared <u>with</u> others.	2	11.76
39	<u>files</u> and so on	2	11.76
40	how to read them <u>fast</u>	2	11.76
41	is a <u>fundamental ability</u> to finish	6	35.29
42	There are many <u>skills</u> to help students	1	5.88
43	Some people think <u>the Reading skill</u>	2	11.76
44	is a <u>correct</u> way to study	4	23.53
45	<u>Then</u> , the listening skill is a	1	5.88
46	Firstly, <u>writing skill</u> is a way to show students' views	3	17.65
48	<u>As everyone know</u> , a paper or an	3	17.65
49	No tutor teaches students <u>only</u> by writing	1	5.88
50	for every <u>foreign</u> student	4	23.53
51	<u>As for</u> speaking, it is a imperative skill	1	5.88
52	be a <u>silent person</u> ?	3	17.65
53	Speaking to others and <u>get</u> more information from others	2	11.76
54	Above all, <u>I think</u> , the most important	2	11.76
55	and a <u>better</u> academic performance.	1	5.88
56	<u>Though</u> reading skill has many advantages	6	35.29
57	Some people think the <u>Reading skill</u> is the most	1	5.88
58	<u>I must</u> use lots	5	29.41
59	use the <u>new</u> knowledge,	2	11.76
60	the content of what proffesser <u>says clearly</u> .	2	11.76
61	way to obtain new <u>things</u> .	10	58.82

62	and <u>indicate</u> how I disagree	3	17.65
63	who finishes <u>amount</u> of tasks and homework,	7	41.18
64	<u>Not only reading skill can</u> help student doing	4	23.53
65	So, what score <u>can a student</u> obtain	1	5.88
66	how to organise words to <u>indicate</u> points	9	52.94
67	to have so many <u>good</u> skills	2	11.76
69	<u>the integrated skill</u> is the best way	1	5.88
70	, the listening skill is a prior skill for <u>a students</u> .	1	5.88
71	As for speaking, it is <u>a</u> imperative skill for every foreign student to own.	4	23.53
72	Speaking skill <u>is</u> communicative skill	1	5.88
73	read them fast and <u>correctly</u>	2	11.76
74	Not only reading skill can help student <u>doing better</u>	1	5.88
75	but <u>also</u> other academic skills	1	5.88
76	but also other academic skills will <u>help more</u>	3	17.65
77	it is a <u>imperative</u> skill for every foreign student to own.	4	23.53
78	students must develop various skills to achieve <u>the</u> goal	2	11.76
86	So, <u>what</u> score can a student obtain	2	11.76
89	There are many skills <u>to</u> help students to study	1	5.88
90	is a <u>very</u> crucial	5	29.41
91	Whatever a student hear and write and <u>how</u> fast a student read, if a student	1	5.88
93	Thus, <u>how</u> to read them fast and	1	5.88
94	<u>No tutor</u> teaches students	1	5.88
95	very effective way to <u>obtain</u> new things	4	23.53
96	<u>Above all</u> , I think, the most important	1	5.88
97	So, what <u>score</u> can a student obtain	1	5.88
98	fundamental ability <u>to finish tasks</u> .	3	17.65
99	that every <u>student</u> wants.	1	5.88
100	There are many <u>skills</u> to help students to study	1	5.88
101	Though reading <u>skill</u> has many advantages.	3	17.65
102	, writing skill is a way to <u>show</u> students' views	3	17.65
103	<u>So</u> , in my opinion, it is the most	2	11.76
104	listening is a <u>very</u> effective way	1	5.88
105	<u>Speaking skill is</u> communicative skill.	2	11.76
106	skill is the <u>best</u> way to obtain	1	5.88
107	I must <u>use</u> lots of knowledge and	3	17.65
108	I must use lots of <u>knowledge</u> and	1	5.88
109	how to organise words to indicate <u>points</u>	2	11.76
111	, <u>a paper or an essay must be finished</u>	2	11.76
113	the listening <u>skill</u>	1	5.88
116	and a <u>better</u> academic performance.	1	5.88
117	think the <u>Reading skill</u> is the most important	1	5.88
120	As a student who finishes amount of <u>tasks</u>	1	5.88
122	Though reading skill <u>has many</u> advantages.	2	11.76
123	the Reading skill is <u>the most important</u> compared with others.	2	11.76
124	<u>own</u> views with others,	1	5.88

125	students must develop <u>various skills</u> to achieve the goal.	1	5.88
126	<u>Such as writing, listening, speaking and so on.</u>	2	11.76
130	to have <u>so many</u> good skills	1	5.88
133	So, in my opinion, <u>it</u> is the most	1	5.88
134	Before you use the new <u>knowledge</u> ,	1	5.88
137	topic and indicate <u>how</u> I disagree	1	5.88
138	However, <u>others hold the opposite view.</u>	2	11.76
139	is a way to show students' <u>views.</u>	1	5.88
140	<u>to master for every student.</u>	1	5.88
143	<u>Speaking to others and get more information from others</u>	1	5.88
144	<u>Thus, how to read them fast</u>	1	5.88
145	<u>So, students must develop various skills</u>	1	5.88
151	<u>show students' views</u>	1	5.88
151	basic quality that every student <u>wants.</u>	1	5.88
151	students <u>must develop</u> various skills	1	5.88
151	need hear the content of <u>what</u> proffesser	1	5.88
	Totals	440	16.81

Appendix 7.14 Table 7.9 Part Two Group A and Group B error categorisations with mode

No	Mode GA	No agree	% agree	Mode GB	No agree	% agree
1	46	13	72	46	6	40
2	9	14	78	9	9	60
3	49	15	83	49	6	40
4	55	12	67	55	7	47
5	27	12	67	27	6	40
6	26	6	33	56	2	13
7	9	17	94	9	8	53
8	27	12	67	27	8	53
9	46	11	61	46	4	27
10	55	10	56	49	4	27
11	54	7	39	54	5	33
12	55	7	39	49	4	27
13	11	11	61	13	5	33
14	55	13	72	55	9	60
15	38	18	100	38	14	93
16	55	9	50	54	5	33
17	55	12	67	55	9	60
18	49	12	67	49	6	40
19	55	17	94	55	7	47
20	49	11	61	49	5	33
21	49	11	61	51	5	33
22	49	9	50	49	5	33
23	55	12	67	55	8	53
24	46	11	61	46	7	47
25	10	13	72	10	12	80
26	11	17	94	11	10	67
27	49	12	67	49	7	47
28	46	15	83	46	8	53
29	55	12	67	55	8	53
30	19	5	28	19	7	47
31	14	15	83	14	12	80
32	46	15	83	46	11	73
33	49	9	50	49	4	27
34	16	5	28	49	7	47
35	16	9	50	54	6	40
36	16	8	44	16	6	40
37	49	9	50	49	7	47
Ave No agr w/mode		12	64		17	47
Ave % agr w/mode		31	64		46	49
Average No Concurrences 23.72						
Average % agreement concurrences 64.11						

Appendix 7.15 Table 7.10 Part Two Group A and B categorisation average confidence scores

No	Error	GA Av per error	GB Av per error
1	To <i>get</i> a perfect academic performan is a basic quality that every student wants.	2.56	2.73
2	To get a perfect academic <i>performan</i> is a basic quality that every student wants.	2.78	2.60
3	perfect academic performan is a <i>basic quality</i> that every student wants.	2.28	2.20
4	<i>Let me discuss the topic</i>	2.17	2.33
5	and indicate how I disagree with the <i>topic</i> .	2.22	2.27
6	Reading <i>skill</i> , undoubtely, is a very crucial factor to master	2.06	1.73
7	Reading skill, <i>undoubtely</i> , is a very crucial factor to master	2.83	2.40
8	Reading skill, undoubtely, is a very crucial <i>factor</i> to master	2.61	2.60
9	As a student who <i>finishes</i> amount of tasks and homework,	2.28	2.20
10	I must use <i>lots of</i> knowledge and documents that I do not know to make it better .	2.44	2.13
11	I must use lots of knowledge and <i>documents that I do not know</i> to make it better .	2.06	2.33
12	I must use lots of knowledge and documents that I do not know to <i>make it better</i> .	2.39	2.27
13	different kinds of books <i>essaies</i> , files and so on.	2.72	2.80
14	different kinds of books <i>essaies</i> , files <i>and so on</i> .	2.22	2.47
15	an essay must be finished <i>in</i> the end of a semester.	2.94	2.93
16	<i>Is it the most important ?</i> I do not think so.	2.17	2.20
17	Is it the most important ? <i>I do not think so</i> .	2.33	2.27
18	So, what score can a student obtain is determined by <i>how to organise words</i> to indicate points	2.11	2.13
19	So, <i>in my opinion</i>	2.39	2.20
20	So, in my opinion , <i>it is the most direct approach</i> to gain great academic performance	2.00	1.87
21	<i>to gain great academic performance</i>	2.28	2.27

22	, the listening skill is <i>a prior skill</i> for a students.	2.11	2.27
23	Before <i>you</i> use the new knowledge, you need hear the content of what proffesser says clearly.	2.39	2.27
24	Before you use the new knowledge, you need <i>hear</i> the content of what proffesser says clearly.	2.50	2.33
25	content of what proffesser says clearly.	2.94	2.80
26	content of what proffesser says clearly.	2.94	2.60
27	No tutor teaches students only by writing on blackboard without <i>any voice</i> .	2.28	2.27
28	As for speaking, it is a imperative skill for every foreign student to <i>own</i> .	2.78	2.60
29	Who <i>wannna</i> be a silent person ?	2.44	2.87
30	<i>Whatever</i> a student hear and write and how fast a student read, if a student can not talk own views with others, it is vein for the student to have so many good skills	2.00	2.07
31	Whatever a student hear and write and how fast a student read, if a student <i>can not</i> talk own views with others, it is vein for the student to have so many good skills	2.61	2.73
32	Whatever a student hear and write and how fast a student read, if a student can not <i>talk</i> own views with others, it is vein for the student to have so many good skills	2.72	2.67
33	Whatever a student hear and write and how fast a student read, if a student can not talk own views with others, <i>it is vein for the student to have so many good skills</i>	2.61	2.00
34	Speaking to others and get more information from others is a correct way <i>to study academy</i> and improve one's ability.	2.17	2.20
35	I think, <i>the most important is not only reading skill</i> . Other skills are not. either .	2.44	2.00
36	I think, the most important is not only reading skill. <i>Other skills are not. either</i> .	2.00	2.20
37	the best way to <i>obtain the biggest</i> progress	2.39	2.13
Averages per Participant		2.41	2.35

Appendix 7.16 Lexical error guidance and framework (NewLEAF3)

Identification

Underline all the lexical errors in the writing (A written lexical error is a lexical form or combination of forms, which would not be made by a writer who has attained IELTS Band 9: the highest score, when writing in the same genre. Include both single word errors (*I was very *locksmith to receive the present.*) and only the erroneous parts of whole phrases (*It was raining *with the dogs and the cat.*).

- 1) Ignore errors with paragraph organization, grammar, punctuation, including capitalisation. (**vienna is *In Austria*). All other errors should be considered lexical.
- 2) Follow the rules below to **exclude grammatical errors from the analysis**:

Grammar Error	Examples	Advice
Article errors	<i>a/an/the/article omission</i>	In fixed phrases, such as 'What's <i>*a matter?</i> ' Here, the error is collocational and is therefore a lexical error.
Inflectional affixes including subject/verb agreement, plurals and tense errors	<i>work/worked/working or dog/dogs</i>	Do include derivational affix errors (e.g. <i>He is very consider*able</i>)
Determiners	<i>Please pass me *this book.</i>	
Infinitive 'to' errors	<i>*Live without love is not *live</i>	
Gerund errors	<i>I like ski*/I like to ski*ing</i>	
Genitive errors (including possession errors)	<i>It's Tom* bike. It is *the bike of Tom.</i>	
Comparative and superlative errors	<i>It is *expensiver, It is the *most big</i>	
Clause errors	<i>It's not difficult *for getting to a hospital and *While waiting, my hamburger went cold</i>	
Relative clause errors (including incorrect relative pronoun and referential errors).	<i>The man *what lives next door is a doctor.</i>	
Countability Errors	<i>Ten items or *less. There are too *much cars.</i>	
Auxiliary verb errors	<i>I *am feel hot,</i>	

- 3) Underline all the LEs in the writing: both single word errors (*I was very *locksmith to receive the present.*) and only the erroneous parts of whole phrases (*It was raining *dogs and the cat*). If possible, underline multiple lexical errors in a phrase separately. However, when a sentence lacks coherence, and it is impossible to separate errors, count them as one per phrase. E.g. **all people there aren't in the globe next a few years*, there are three, based on the three phrases (noun phrase, verb phrase and prepositional phrase).
- 4) Count each repeated error.
- 5) Ensure you include instances of infelicitous lexis, including errors of tautological expression (e.g. **very crucial*), word order in common expressions only, errors of variety, (e.g. American English), errors of political correctness, genre errors (especially formality errors), but exclude language that you do not necessarily consider to be wrong, but where you have a preference for alternative expression.
- 6) Read the text twice for errors, as many are easily missed. Refer to the guidance during the analysis.

Categorisation

- 1) If the meaning is unclear after re-reading original composition, allocate it to coherence category.
- 2) Categorise the error based on what the student wrote, not what they should have written. E.g. If a student writes *Oranges taste *not nice*, meaning, Oranges taste awful, this would be a B3 phrase misselection error, code 37, not a code 25 B2D adverb error. The student made an error with 'not nice', not the missing adverb.
- 3) Avoid speculation of what caused the error.
- 4) If errors could be categorised in more than one way, allocate to the more specific/descriptive/helpful category. E.g. if a suffix or prefix contains a spelling error, allocate to suffix or pre-fix. If there is a phrase error with informal language, allocate to Formality.

Omission = omitted letters/words are required
 Over-inclusion = extra letters/words should not be there
 Mis-selection = wrong letters/words chosen
 Mis-order = letters/words in the wrong order

Error type	Description/advice	Examples	OMISSION	OVER-INCLUSION	MIS-SELECTION	MIS-ORDER
Section A Form						
A1 SUFFIX	Correct root, but incorrect/missing/misspelt suffix. Do not include verb tense errors (<i>He was go*</i>) subject/verb agreement errors (<i>He *go</i>) or errors with plurals.	<i>They were very consider*able</i> <considerate> <i>The data were analys*t.</i> <analysed>	1	2	3	N/A
A2 PREFIX	Correct root, but incorrect/missing/misspelt prefix.	<i>It's quite *inimportant.</i> <unimportant>	4	5	5	N/A
A3 SPELLING (misc)	Meaning of intended word is clear and exists in English. Words with unclear meaning should be categorised as B1 .	<i>I need to *safe some money.</i> <save> <i>Put the *folwers in water.</i> <flowers>	6	7	8	9
A4 TWO WORDS SHOULD BE ONE	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-law or ice cream/ ice-cream). Does not include an extra incorrect word (see Section B2).	<i>*Every one will pay.</i> <everyone> <i>I found it *on line.</i> <online>	10			
A5 ONE WORD	Ignore debatable points and hyphenated/non hyphenated words (mother in law/mother-in-	<i>Where is the *paperbin?</i> <paper bin> <i>Put it on the *dinnertable.</i> <dinner table>	11			

SHOULD BE TWO	law or ice cream/ ice-cream). Does not include missing word (see Section B2).					
Section B Meaning			OMISSION	OVER-INCLUSION	MIS-SELECTION	MIS-ORDER
B1 COHERENCE	Words do not exist in English or meaning of whole phrase is unclear. Only select this category if you have re-read and still cannot be sure of the meaning of the word or phrase.	<i>*Where is those that are? <?></i> <i>I am *lipsh <?></i>	12			
Errors with Single Words (if clearly a misspelling, allocate to A3, e.g. I need to *safe some money. <save>)						
B2A CONJUNCTION	Include all conjunctions or transition signals here, including multiword items, such as 'in spite of '.	<i>It was hot *and I took off my coat. <so></i> <i>In *despite of the rain, we went swimming. <spite></i>	13	14	15	16
B2B NOUN	Include compound nouns, but not multi-word units/collocations/fixed expressions.	<i>It was hot so I took off my *hand. <coat?></i> <i>We visited the *art museum. <art gallery></i>	17	18	19	N/A
B2C ADJECTIVE		<i>I feel *serious. <stressed></i> <i>The views were *handsome. <beautiful></i>	20	21	22	23
B2D ADVERB		<i>He sings *deliciously.</i> <i>They are growing *quick. <quickly></i>	N/A	24	25	N/A
B2E PREPOSITION	Include prepositions of time and place and dependent propositions, not errors with infinitive 'to'.	<i>I woke up *on 6am. <at></i> <i>See you *in the corner. <at/on></i> <i>I was suspicious *about his behaviour. <of></i>	26	27	28	N/A

	Include incorrect phrasal verb errors with particles/prepositions here.					
B2F PRONOUN		<i>My mother, *she is ill.</i>	29	30	31	32
B2G VERB	Include incorrect phrasal verb errors with roots verbs here.	<i>Do you *get the time, please? <have> He *put on the table. <laid></i>	33	34	35	36
Errors that span more than one word			OMISSION	OVER-INCLUSION	MIS-SELECTION	MIS-ORDER
B3 PHRASE ERROR (MISC)	Meaning is clear, but word choices or word ordering is non-expert-like. There may be more than one error in the phrase (or multi-word unit). Also use for whole problematic phrases which require re-writing.	<i>* The number of people is really great. <There is a large number of people>. *Next a few years. <In the next few years.> Oranges taste *not nice,<awful></i>	N/A	N/A	37	38
B3A MIS-ORDERING IN FIXED PHRASES	Collocational errors relating to order of pairs of words or groups of three.	<i>*hike-hitch <hitch-hike> *fro and to <to and fro> *Dark, tall and handsome <Tall, dark and handsome></i>	39			
B3B VERBOSITY	Too many words used or points are repeated unnecessarily. Categorise single extra word in B2 overinclusion above.	<i>He bought an apple and* he bought a banana. <He bought an apple and a banana.> I like cake. *I like cake because it is sweet. <I like cake because it is sweet.></i>	40			

B3C UNDERSPE CIFICATION	Phrases/sentences that require more detail to improve expression. Categorise single missing word in B2 omission above. If meaning is unclear, categorise as B1 .	<i>Although *cars in the country are lower...</i> <Although there are lower car numbers in the country....> <i>We cut *because tree high.</i> <We cut the trees because they were getting too high.>	41
B3D FORMALITY	Phrases that are correct but seem overly formal or informal for the genre. If there is a formality error with a single word, allocate here.	<i>I *informed my girlfriend via the medium of the telephone.</i> <I told my girlfriend on the phone.> <i>*Milk stocks in the refrigerator are starting to ebb.</i> <We need more milk.>	42
B4 UNCATEGO RISABLE	Only use this category after considered thought and if the error cannot fit into any category above.	?	43

LEA Flowchart

Complete the analysis by following the steps below.

1) Collection of scripts

No recourse to dictionary or other reference materials, including internet resources. Exam conditions. If using MS Word, turn off the grammar and spell checkers.

2) Preparation of script

Type the student writing out into Microsoft word. Save as a word document.

3) Identification of errors

An error is 'a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speaker's highly-skilled user of English counterparts'. (Lennon 1991), 'unless the error is a norm of that variety of English and made in that country' (George 1972).

Identify all errors in the writing. Include writing that seems 'odd or non-expert-like', depending on context (for example, consider whether the writing would be penalised if it is in an IELTS context).

Identify the domain and extent of the error:

Domain = The breadth of context which the analyst must examine to understand the error.

Extent = The span of the utterance, or actual words, which must be changed in order to fix the error.

Optional stage: Count the number of clauses that are error free and the number of clauses that contain errors to come up with an accuracy ratio.

4) Identification of grammatical and lexical errors

Use the following criteria to identify the grammatical errors

- I. Article errors (e.g. a/an/the/no article), except in fixed phrases, such as '*What's *a matter?*' Here, the error is collocational and is therefore a lexical phrase error.
- II. Inflectional affixes (e.g. work/worked/working or dog/dogs). Do include derivational affix errors (e.g. *He is very consider*able*).
- III. Errors with infinitive 'to' (e.g. **Live without love is not *live*)
- IV. Gerund errors (e.g. *I like ski**.)
- V. Genitive errors including errors used to show possession. (e.g. *It's Tom* bike, It is *the bike of Tom.*)
- VI. Comparative and superlative affixes (e.g. *It is *expensiver, It is the *most big*).
- VII. Clause errors (e.g. *It's not difficult *for getting to a hospital*)
- VIII. Errors in relative clauses including incorrect relative pronoun and referential errors. (e.g. *The man *what lives next door is a doctor*)

- IX. Errors in countability (*Ten items or *less.*)
- X. Errors in tense (*Yesterday I *go to school.*)

Ignore errors with punctuation, including capitalisation. (**vienna is *In Austria*).

All other errors are lexical

5) Separating Lexical Errors

Multiple lexical errors in a phrase are counted separately. However, when a sentence lacks coherence, and it is difficult to separate errors, count them as one per phrase. For example, in the sentence *“*all people there aren't in the globe next a few years’*, there are three, based on the three phrases (noun, verb and prepositional phrase).

When counting errors, only include one instance of the same error.

6) Categorisation of lexical errors

Process

1. Refer to the framework to categorise lexical errors. Consider surface taxonomy (misselection, omission, overinclusion, misorder) and type of error (form or meaning)
2. Decide if the error is an error of form (spelling, spacing or error with prefix or suffix)
3. Decide if the error is limited to one word or the whole phrase requires rewriting.
4. If it is one word that can be replaced, allocate to one word/word class/surface taxonomy.
5. If the whole phrase requires rewriting, allocate to phrase and most appropriate sub category
6. Where meaning of the error is unclear, consult the full original context and co-text to help you understand the intended meaning of the error. Then try to make an expert-like, plausible reconstruction of the error.
7. Categorise the error that was made, not the type of reconstruction required to fix it.
8. If the meaning of the error cannot be understood, allocate to the coherence category.
9. Only if the error cannot be justifiably categorised in A1-B4, allocate to the uncategorizable category.

Advice

- Avoid speculation of what caused the error at this stage, as this will result in less reliable results. Simply categorise the actual error made. Causality comes later.

Categorisation can be done by using the track changes feature in MS word in the digitised script or errors can be recorded in the framework itself by either copying and pasting the errors into the corresponding cells if only using one script or if using scripts from a cohort, tally marks can be recorded in the framework.

Identification of cause of errors

If you are bilingual, you may be able to categorise the first language transfer errors into

- Calque (translation word by word)
- Coinage (a new word or phrase is created by the student, based on their first language)
- Borrowing (a word taken from another language)

If you are the teacher and familiar with the teaching context, you may be able to identify errors caused by

- materials
- teacher explanation
- or rule 'over application'.

Identify the more 'serious' errors

To decide on which errors to give feedback to your students on, if you are able, further sub-divide the errors into

- Coherence Errors (those which cause confusion). These are the more serious.
- Irritating errors (those that cause a degree of irritation).
- Mistakes (those errors with language that has been previously taught) Vs Pre-systemic errors with language that has not yet been taught) Vs Slips errors that you know the student will be able to identify as such by themselves).
- Basic errors (those that are made by students whose general proficiency level states that they should not be making this error).
- Errors made with lexis that may be considered above the proficiency level of the student.
- Completely wrong errors, Infelicitous errors and errors a highly-skilled user of English might make
- Collocational errors.