L2 Decoding and reading comprehension in Saudi foundation year ESL students: An investigation into readers' decoding abilities, outputs, experiences, and processes, and the contribution of their L2 decoding proficiency to their L2 reading comprehension

Hala Alghamdi

A thesis submitted for the Degree of Doctor of Philosophy in Applied Linguistics

Word count: 91,965 excluding preliminary pages, references, and appendices

St. Stephen's House

Trinity term, 2020

Table of Contents

Table of Contents 2
Index of Tables 8
Index of figures 10
List of Abbreviations 12
Thesis Abstract 14
1. Introduction 15
1.1. Research into second language decoding: The decoding abilities, realisations, experiences and processes of second language readers 15
1.2. Research into the contribution of L2 decoding proficiency to L2 reading comprehension 18
1.3. The Saudi setting for ESL reading in Higher Education 20
1.4. The aims of the present study: 21
1.5. The structure of the thesis 22
2. Literature Review 25
2.1. Models of reading comprhension: 25
2.1.1. Process models of reading comprehension 26
2.1.2. Componential approaches of reading comprehension 27
2.1.2.1. Review of existing componential models: 28
2.1.2.1.1. The Simple View of Reading 28
2.1.2.1.2. Bernhardt's componential model of L2 reading comprehension 31
2.1.2.2. Components (predictors) of L2 reading comprehension ability 33
2.1.2.2.1. L2 Vocabulary knowledge (L2VK) 34
2.1.2.2.2. L2 Grammar knowledge (L2GK) 34
2.1.2.2.3. L1 reading comprehension (L1RC) 35
2.1.2.2.4. L2 listening comprehension (L2LC) 35
2.1.2.2.5. Other components 36
2.1.2.2.6. Decoding proficiency 37
2.2. Second language decoding acquisition 52
2.2.1. Writing systems: an overview 52
2.2.2. L1 writing systems and L1 decoding 53
2.2.3. L2 decoding acquisition: theoretical framework 56
2.2.3.1. Orthographic Distance Hypothesis (Koda, 2008; Koda, 2013) 56
2.2.4. The Arabic writing system 57
2.2.5. Empirical evidence 59
2.2.5.1. Quasi-experimental studies 60
2.2.5.1.1. L2 decoders with congruent versus incongruent L1WS backgrounds 60
2.2.5.1.2. Quasi-experimental studies: conclusion and remarks 65
2.2.5.2. Descriptive studies 67

2.2.5.2.1. Descriptive studies: conclusion and remarks 71
2.2.5.3. Qualitative studies 72
2.2.5.3.1. Qualitative studies: conclusion and remarks 74
2.2.5.4. Conclusion and gaps in previous research into the decoding challenges of second language readers 75
2.3. The present study: Aims and research questions 76
3. Methodology 77
3.1. Research Design 77
3.2. Sampling 82
3.2.1. Population 82
3.2.2. Selection of the Higher Education Institution 82
3.2.3. Participants' selection criteria 83
3.2.4. Sample size 84
3.2.5. Sampling strategy 85
3.3. Overall Procedure 86
3.4. Data collection methods 89
3.4.1. Quantitative data collection methods 89
3.4.1.1. Measures 89
3.4.1.1.1. L2 decoding proficiency 89
3.4.1.1.2. L2 reading comprehension 94
3.4.1.1.3. L2 vocabulary knowledge 96
3.4.1.1.4. L2 grammar knowledge 98
3.4.1.1.5. L1 reading comprehension 99
3.4.1.1.6. Background questionnaire 100
3.4.1.2. Summary of the tests 101
3.4.2. Qualitative data collection methods 102
3.4.2.1. Audio-recordings of learners' decoding outputs 102
3.4.2.1.1. Inter- and intra- transcriber reliability 104
3.4.2.2. Interview A: The decoding experiences of Saudi foundation year ESL students 105
3.4.2.3. Think-aloud reports 106
3.4.2.4. Interview B: The contribution of L2 decoding proficiency to L2 reading comprehension: readers' experiences 109
3.5. Data analysis methods: 111
3.5.1. Part 1: The decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students 111
3.5.1.1. The decoding abilities of Saudi foundation year ESL students 112
3.5.1.2. The decoding realisations of Saudi foundation year ESL students 112
3.5.1.3. The decoding experiences of Saudi foundation year ESL students 117
3.5.1.4. The decoding processes of Saudi foundation year ESL students 117

3.5.2. Part 2: The contribution of L2 decoding proficiency to L2 reading comprehension 117
3.5.2.1. The relationships between L2 reading comprehension performance, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 decoding proficiency, L2 real word decoding proficiency, and L2 pseudoword decoding proficiency 118
3.5.2.2. The contribution of L2 decoding proficiency *before and after* controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability 119
3.5.2.3. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner 121
3.5.2.4. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the decoding measure 123
3.5.2.5. The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance Analysis and Relative Weight Analysis 124
- Dominance Analysis 127
- Relative Weight Analysis 128
- Dominance Analysis or Relative Weight Analysis? 129

3.5.2.6. The contribution of L2 decoding proficiency to L2 reading comprehension performance: qualitative insights from readers' experiences 131
3.6. The piloting phase 131
3.7. Ethical considerations 134
4. Results I: The decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students (CEFR: A1, A2, and B1) 137
4.1. The decoding abilities of Saudi foundation year ESL students (CEFR: A1, A2, and B1). 137
4.1.1. Descriptive statistics and ANOVA 137
4.2. The decoding realisations of Saudi foundation ESL year students 139
4.2.1. Categories of decoding errors 139
4.2.1.1. Vowel-related errors 140
4.2.1.1.1. Short vowel phonemes 141
4.2.1.1.2. Long vowel phonemes 151
4.2.1.1.3. Diphthongs 159
4.2.1.2. Consonant-related errors 163
4.2.1.3. Phoneme insertion, omission, and mis-ordering errors 167
4.2.1.4. Whole-word errors 169
4.2.2. Conclusion: 170
4.3. The decoding experiences of Saudi foundation year ESL students 172
4.3.1. Results: 172
4.3.1.1. Lack of L2 GPC knowledge 172
4.3.1.2. Word Length 174

4.3.1.3. Unfamiliar words 176
4.3.1.4. Silent letters 177
4.3.1.5. English Script 178
4.3.1.6. Other 179
4.3.2. Conclusion 180
4.4. The decoding processes of Saudi foundation year ESL students 181
4.4.1. Results: 181
4.4.1.1. Automatic processing 181
4.4.1.2. Assembled processing 183
4.4.1.2.1. Dividing words into units 183
4.4.1.2.2. Conscious use of L2 GPC rules 188
4.4.1.3. Other: Partial (incomplete) processing 194
4.4.2. Conclusion 195
5. Results II: The contribution of second language decoding proficiency to second langauge reading comprehension 196
5.1. Data screening and preliminary analyses 196
5.1.1. Data screening: 196
5.1.1.1. Accuracy of input 196
5.1.1.2. Checking for missing data 196
5.1.1.3. Checking for univariate outliers 197
5.1.1.4. Checking for multivariate outliers 198
5.1.1.5. Checking for normality 201
5.1.2. Descriptive statistics 203
5.1.2.1. Descriptive statistics for the background questionnaire 203
5.1.2.2. Descriptive statistics for L2 reading comprehension, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, and overall L2 decoding proficiency measures 205
5.2. Answering the research question: What is the contribution of second language decoding proficiency to second language reading comprehension performance? 206
5.2.1. The relationships between L2 reading comprehension performance, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 decoding proficiency, L2 real word decoding proficiency, and L2 pseudoword decoding proficiency 207
5.2.2. The contribution of overall L2 decoding proficiency to L2 reading comprehension *before* and *after* controlling for L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability 210
5.2.3. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner 215

5.2.4. The contribution of second language decoding proficiency to second language reading comprehension as a function of the decoding measure 218
- Analysis A: 218
- Analysis B: 221

5.2.5. The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance analysis and Relative weight analysis 226
- The relative importance of overall L2 decoding proficiency to L2 reading comprehension 226
- The relative importance of L2 real word and L2 pseudoword decoding proficiency to L2 reading comprehension 228

5.2.6. The contribution of second language decoding proficiency to second language reading comprehension: qualitative insights from readers' experiences 232
5.2.6.1. L2 Vocabulary knowledge 232
5.2.6.2. Affective factors 233
5.2.6.3. Time pressure 234
5.2.6.4. L2 Grammar knowledge 234
5.2.6.5. Background/topic knowledge 235
5.2.6.6. Inferencing ability 235
5.2.6.7. Other 235
5.2.6.8. L2 Decoding knowledge 236

6. Discussion 242

6.1. The decoding abilities, realisations, experiences, and processes of Saudi foundation year students 242
6.1.1. Decoding abilities: 242
6.1.1.1. The findings on Saudi L2 learners' decoding abilities and the literature 242
6.1.1.2. Implications for theories of reading 243
6.1.1.3. Implications for language pedagogy 244
6.1.1.4. Limitations of the findings 245
6.1.2. Decoding realisations and experiences 246
6.1.2.1. Findings on learners' decoding realisations and experiences 246
6.1.2.1.1. Decoding realisations: 246
6.1.2.1.2. Decoding experiences 248
6.1.2.2. Implications for theories of reading 248
6.1.2.3. Implications for language pedagogy 249
6.1.2.4. Limitations of the findings 250
6.1.3. Decoding processes 250
6.1.3.1. The findings on learners' decoding processes and the literature 250
6.1.3.2. Implications for theories of reading 253

6.1.3.3. Implications for language pedagogy 255
6.1.3.4. Limitations of the findings 256
6.2. The contribution of second language decoding proficiency to second language reading comprehension 257
6.2.1. The findings on the contribution of L2 decoding proficiency to L2 reading comprehension and the literature 257
6.2.1.1. The Contribution of L2 decoding proficiency to L2 reading comprehension 257
6.2.1.2. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner 259
6.2.1.3. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the L2 decoding measure 260
6.2.1.4. The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance analysis and Relative weight analysis 262
6.2.1.5. The contribution of L2 decoding proficiency to L2 reading comprehension performance: qualitative insights 263
6.2.2. Implications for theories of reading 265
6.2.3. Implications for language pedagogy 266
6.2.4. Limitations of the findings 267
7. Conclusion 269
7.1. Recommendations for future research 271
7.1.1. Documenting the decoding abilities of second language learners 271
7.1.2. The nature of second language of decoding development 272
7.1.3. Validity and reliability studies into the use of think-aloud protocols as means for investigating word decoding processes 273
7.1.4. Structural equation modelling studies into the contribution of second language decoding proficiency to second language reading comprehension 273
7.1.5. A meta-analysis of the contribution of L2 decoding proficiency to L2 reading comprehension and the factors that may moderate this contribution. 274
7.1.6. Longitudinal studies into the contribution of L2 decoding proficiency to L2 reading comprehension performance 274
7.1.7. The contribution (or lack thereof) of L2 decoding proficiency to L2 listening comprehension performance 275
References 276

Index of Tables

Table 1. Illustration of the basic mapping principles in alphabetic English, consonantal Arabic, syllabic Japanese (Kana), and logographic Chinese for the word ‘cat’. 52
Table 2. Outline of the data collection sessions.....88
Table 3. A sample passage from Gray Silent Reading Test (Form A) 96
Table 4. Summary of measures.....102
Table 5. An example illustrating some of the problems associated with Beta weights as a measure of relative importance in the context of correlated predictors (adapted from Tonidandel and LeBreton (2011).....124
Table 6. An example illustrating how General Dominance Weights are calculated (adapted from Azen and Budescu (2003))....128
Table 7. Facility level of short vowel phonemes 142
Table 8. A breakdown of the realisations of /i/ across two graphemic contexts (<i> and <e>) 144
Table 9. A breakdown of the realisations of /ə/ across five graphemic contexts (<ou>, <o>, <er>, <e>and <->) 149
Table 10. A summary of the short vowel phonemes and their three most frequently observed realisations 150
Table 11. Facility level of long vowel phonemes 152
Table 12. A breakdown of the realisations of /u:/ across three graphemic contexts (<oo>, <ew>, and <oe>).....154
Table 13. A breakdown of the realisations of /ɔ:/ across two graphemic contexts (<a> and <o>).....155
Table 14. A breakdown of the realisations of /ɜ:/ across four graphemic contexts (<ur>, <ir> ,<er>, and <or>).....157
Table 15. A summary of the long vowel phonemes and their three most frequently observed realisations 158
Table 16. Facility level of diphthongs 159
Table 17. A summary of the target diphthongs and their three most frequently observed realisations 162
Table 18. Facility level of consonant phonemes.....164
Table 19. The realisations observed for consonant phonemes along with their percentage frequency.....166

Table 20. Summary of emerging categories 172
Table 21. Amount of missing data on each of the continuous variables 196
Table 22. Mahalanobis distance, Cook's distance, standardised residuals, Studentised residuals, and Leverage values for cases 111 and 137.....200
Table 23. Skewness and kurtosis z scores of all the study variables. 202
Table 24. Skewness and kurtosis values and z scores prior transformation of the L2 vocabulary knowledge measures, post square root transformation, and post logarithm transformation 203
Table 25. Descriptive statistics for the background questionnaire 204
Table 26. Descriptive statistics for the study variables.....205
Table 27. Spearman correlations between study variables. 208
Table 28. Order of variable entry.....210
Table 29. Summary of hierarchical regression analysis for variables predicting L2 reading comprehension 214
Table 30. Summary of hierarchical moderated regression analysis 216
Table 31. Summary of follow up simple slope analysis 217
Table 32. Summary of regression analyses for investigating the contribution of L2 pseudoword decoding proficiency to L2 reading comprehension 219
Table 33. Summary of regression analyses for investigating the contribution of L2 real word decoding proficiency to L2 reading comprehension.....222
Table 34. Summary of follow up simple slope analysis 224
Table 35. The relative importance of overall L2 decoding proficiency to L2 reading comprehension using Dominance, Relative, and Beta weights 227
Table 36. The relative importance of real word and pseudoword decoding proficiency to L2 reading comprehension using Dominance, Relative, and Beta weights 228

Index of figures

Figure 1. The basic idea underlying the Universal Grammar of Reading (From Perfetti, 2003) ..... **Error! Bookmark not defined.**
Figure 2. Standard English vowel chart 59
Figure 3. Study design (part 1) 80
Figure 4. Study design (part 2) 81
Figure 5. A sample VLT task 96
Figure 6. A sample grammar task from Shiotsu (2010) 98
Figure 7. Interview 2 outline 111
Figure 8. A summary of the analytical procedures applied to analyse learners' decoding realisations 116
Figure 9. Order of variable entry in hierarchical regression 120
Figure 10. Order of variable entry in hierarchical regression 2 121
Figure 11. Categories of decoding errors in the whole-sample analysis along with their percentage frequency (of the total number of errors observed). 139
Figure 12. Categories of decoding errors in the subsample analysis along with their percentage frequency (of the total number of errors observed) 140
Figure 13. Realisations of the vowel phoneme /e/ and their percentage frequency 143
Figure 14. Realisations of the vowel phoneme /i/ and their percentage frequency. 144
Figure 15. Realisations of the vowel phoneme /ʌ/ and their percentage frequency. 145
Figure 16. Realisations of the vowel phoneme /ʊ/ and their percentage frequency 146
Figure 17. Realisations of the vowel phoneme /ɒ/ and their percentage frequency. 147
Figure 18. Realisations of the vowel phoneme /æ/ and their percentage frequency. . 148
Figure 19. Realisations of the vowel phoneme /ə/ and their percentage frequency. 149
Figure 20. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target vowel /e/ (circled in blue). Note: */eː/ does not exist in RP 151
Figure 21. Realisations of the vowel phoneme /aː/ and their percentage frequency. 153
Figure 22. Realisations of the vowel phoneme /iː/ and their percentage frequency. 153
Figure 23. Realisations of the vowel phoneme /uː/ and their percentage frequency. 154
Figure 24. Realisations of the vowel phoneme /ɔː/ and their percentage frequency. 155
Figure 25. Realisations of the vowel phoneme /ɜː/ and their percentage frequency. 156

Figure 26. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target long vowel /i:/ (circled in blue). (note: the arrow indicates the diphthong /ei/).....158
Figure 27. Realisations of the vowel phoneme /aɪ/ and their percentage frequency 160
Figure 28. Realisations of the vowel phoneme /eɪ/ and their percentage frequency. 160
Figure 29. Realisations of the vowel phoneme /əʊ/ and their percentage frequency. 161
Figure 30. Realisations of the vowel phoneme /aʊ/ and their percentage frequency. 162
Figure 31. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target diphthongl /ai/ (circled in blue). 163
Figure 32. Cook's distance of the 214 cases 199
Figure 33. Histogram of normality of residuals.....211
Figure 34. Normal p-p plot of regression standardised residuals 211
Figure 35. Scatterplot of residuals against predicted reading comprehension scores.....212
Figure 36. The contribution of overall L2 decoding proficiency as a function of the proficiency level of the learner 217
Figure 37. The contribution of L2 real word decoding proficiency as a function of the proficiency level of the learner 224

List of Abbreviations

CI	Confidence intervals
DP	Decoding proficiency
DV	Dependent variable
ESL	English as a second language
ELI	English Language Institute
EMI	English-medium instruction
ETS	Educational Testing Service
FPT	Full pphonological transcriptions
GK	Grammar knowledge
GPC	Grapheme-phoneme correspondence
IELTS	International English Language Testing System
HE	Higher Education
HEI	Higher Education Institution
IV	Independent variable
L1	First language
L2	Second language
L2DP	L2 decoding proficiency
L2GK	L2 grammar knowledge
L2PDP	L2 pseudoword decoding proficiency
L2RDP	L2 real word decoding proficieny
L1RC	L1 reading comprehension ability
L2RC	L2 reading comprehension
L2VK	L2 vocabulary knowledge
L1WS	First language writing system
L2WS	Second language writing system
MD	Mahalanobis Distance
RQ	Research question
RC	Reading comprehension
SLA	Second language acquisition
SPT	Subsample of the phonological transcriptions
TOEFL	Test of English as a Foreign Language

TOWRE	Test of Oral Word Reading Efficiency
VK	Vocabulary knowledge
VIF	Variance inflation factor
WS	Writing system

Thesis Abstract

Phonological decoding proficiency has attracted ample attention in L1 reading research, and has been a key variable in prediction studies of L1 reading comprehension ability. Contrary to L1 research, L2 decoding skill has received less attention, with various linguistic contexts remaining unexplored (e.g., Arabic L1 learners of alphabetic languages). Furthermore, the unique contribution, or lack thereof, of this skill to L2 reading comprehension is unclear. Consequently, this study aimed to: (1) explore L2 decoding among Saudi (Arabic L1) university ESL students (CEFR: A1, A2, B1), particularly, the extent to which they are able to perform on measures of English decoding, and the nature of their decoding outputs, experiences, and processes; (2) investigate the contribution of L2 decoding proficiency to L2 reading comprehension, relative to the contributions of some other variables known to be important predictors of this skill: namely, L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension ability.

To achieve the first aim, 216 Saudi foundation year ESL students took part in a real and pseudoword reading-aloud task aimed at assessing their decoding proficiency. The decoding outputs of a subsample (N=60) were then phonemically transcribed and analysed. A qualitative investigation was also carried out in which further subsamples took part in: (a) an interview aimed at understanding their decoding experiences (N=12); and (b) a think-aloud task aimed at exploring their decoding processes (N=15).

To achieve the second aim, the following additional measures were administered to the same 216 participants: L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension ability. A complementary qualitative investigation was also carried out in which 18 participants took part in interviews aimed at validating and providing greater depth to the quantitative results.

With respect to the first aim, descriptive analyses suggest that students encounter difficulties with English decoding, with a mean overall decoding percentage of success of 35%, 44%, and 52% for A1, A2, and B1 students, respectively. Error analyses of their decoding outputs and thematic analyses of their self-reported decoding experiences highlighted the following: difficulties with decoding vowel, consonant, and silent graphemes; a tendency to add to, omit, and mis-order existing graphemes; a tendency to make whole-word substitutions; difficulties with decoding unfamiliar and long words, and difficulties with the L2 script. Thematic analysis of the think-aloud data showed a range of decoding processes, including automatic processing, small-unit processing, large-unit processing, analogy-guided processing, intuition-guided processing and partial processing.

With respect to the second aim, a series of regression analyses showed that: (a) there was a significant relation between L2 decoding proficiency and L2 reading comprehension performance; (b) the contribution of L2 decoding proficiency to L2 reading comprehension was significant before and also after controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension; (c) L2 decoding proficiency was the third strongest predictor of L2 reading comprehension after L2 grammar knowledge and L2 vocabulary knowledge; (d) the two decoding measures (real word and pseudoword decoding) appeared to be more related than different in that neither made a significant contribution to L2 reading comprehension after controlling for the other and for other measured covariates; (d) the contribution of L2 decoding proficiency seemed to decrease as linguistic proficiency increased. Further, a thematic analysis of the interview data suggested that decoding difficulties could interfere with learners' comprehension, with this being due to the role decoding plays in facilitating meaning access, modulating situational affect, and possibly freeing cognitive capacity. Together, these findings have various implications for reading theory, pedagogy, and assessment.

1. Introduction

Decoding Proficiency (DP), the ability to transfer written codes to phonological codes accurately and quickly, has attracted ample attention in first language (L1) reading research. This can be seen in the large number of decoding studies across a range of educational and writing-system contexts (see Ziegler & Goswami, 2005). In parallel with these investigations, another major line of L1 reading research, componential (prediction) studies of reading comprehension, has regarded decoding proficiency an essential component (predictor) of reading comprehension, repeatedly investigating its power in predicting L1 reading comprehension performance (e.g., Cunningham, Stanovich, & Wilson, 1990).

Contrary to L1 researchers, L2 researchers have given L2 decoding ability little attention (Koda, 2005), and still less attention to its contribution to the ultimate goal of decoding: reading comprehension (Nassaji, 2014; Pasquarella, Gottardo, & Grant, 2012). In what follows, these two gaps are outlined broadly, and with reference to the L1 writing system (WS) of the participants of the present study (Arabic L1). These are followed by a description of the setting of the present research (Higher Education in Saudi Arabia). Finally, the aims of the present study are presented, followed by an outline of the structure of this thesis.

1.1. Research into second language decoding: The decoding abilities, realisations, experiences and processes of second language readers

In the second language acquisition context, decoding skill has received less attention than it deserves, presumably the consequence of the assumption that L2 decoding will eventually develop alongside L2 linguistic proficiency, traditionally viewed mainly as L2 vocabulary and grammar knowledge (Koda, 2005). However, as Koda (2005) argued, there is evidence for the prevalence of inefficient decoding among fluent bilinguals, suggesting that linguistic knowledge, while necessary for skilled reading, does not presume proficient decoding, and that acquisition of L2 decoding ability should not be readily assumed. This lack of attention to L2 – hereafter English – decoding skill persists despite growing evidence for links between L2 decoding ability and L2 word learning (Hamada & Koda, 2010), motivation to learn the second language (Erler & Macaro, 2011), and, potentially, L2 reading comprehension (Jiang, Sawaki, & Sabatini, 2012).

Furthermore, of the L2 decoding studies available, the population of focus has often been L2 students in native English settings, for example, USA (Hamada and Koda, 2010), and particularly those from logographic L1 writing system backgrounds (for example, Chinese) (Cook & Bassetti, 2005), leaving a large area to be investigated. Most notably, much is yet to

be known about L2 decoding in readers with non-logographic L1 writing system backgrounds. Previous research suggests that L2 readers, particularly those with reasonably well-developed L1 literacy skills and still-developing L2 literacy skills, have developed decoding processes in response to their L1 literacy experience (e.g., Hamada & Koda, 2008). Of late, L2 reading researchers have noted that L1 decoding processes remain active during the L2 decoding process. Further, when these L1 decoding processes are dissimilar to those of the L2, in response to differences in the writing system, L1 decoding processes will need to undergo major adjustments to be serviceable in the L2 writing system (e.g., Hamada & Koda, 2008). This prior processing experience, while facilitative in some respects, has been argued to pose challenges to the L2 decoder, as evident in research examining L2 learners' decoding abilities, realisations (outputs), processes, and experiences (e.g., see Hamada & Koda, 2008; Woore, 2016, 2010; and Erler, 2003).

Given the influence of L1WS experience on L2 decoding, it is essential for L2 decoding research to examine learners from a wide variety of L1WS backgrounds (for both theoretical and pedagogical reasons, as will be discussed in detail later). Unfortunately, to date, this is far from being the case. Specifically, L1 Arabic readers are largely underrepresented in L2 decoding research. These readers remain under-represented though they present an interesting case. Typically, Arabic is considered a consonantal writing system. This consonantal WS differs from alphabetic English in a number of ways: its representational unit (primarily consonantal phonemes vs. English consonantal and vowel phonemes); its orthographic depth (almost one-to-one consonantal GPC vs. English relatively many-to-one GPC); and its physical properties (right-to-left directionality and non-Roman script vs. English left-to-right directionality and Roman script), all of which could potentially pose difficulties for the L2 decoder. Furthermore, the picture is made more complex by the fact that Arabic readers are taught to read an alphabetic L1WS in their first three years of schooling (before progressing to read the consonantal writing system) (Abu-Rabia & Taha, 2006). This early years' writing system is alphabetic, but is still different from English. It exhibits different levels of orthographic depth (almost one-to-one GPC vs. English relatively many- to- one GPC) as well as different physical properties (i.e., again right-to-left directionality and non-Roman script vs. English left-to-right directionality and Roman script). Thus, in a sense, Arabic readers of English are literate in two writing systems: alphabetic and consonantal. This rich L1WS background makes it difficult to make assumptions about their L2 decoding challenges based on simple contrastive L1WS-L2WS analysis, thus further highlighting the need to directly study this group of L2 learners.

Apart from limitations in the L2 contexts surveyed, and as will be discussed in detail later, prior L2 decoding research has been limited by a number of methodological issues. These include: (a) the use of idiosyncratic, rather than widely recognized, decoding measures when assessing L2 decoding skill. This practice might not only hinder valid and reliable assessment of learners' decoding abilities, but might also make it difficult to properly quantify the decoding ability of L2 readers throughout the course of L2 development; (b) a tendency to make a priori assumptions of the decoding difficulties of learners based on contrastive L1-L2WS analysis, though these assumptions may be incomplete and/or, worse, inaccurate; (c) small-scale analyses of learners' decoding realisations, though thorough error-analyses can provide further insights into the decoding difficulties of L2 readers; (d) employing single research methods to studying the decoding challenges of L2 learners. For instance, examining learners' decoding abilities (by means of their performance on decoding tests), or their decoding realsiations, experiences or processes, though incorporating these lines of research in a single study could provide a fuller understanding of the decoding challenges of L2 readers; (e) Employing learners from a single level of proficiency, thus making it difficult to see whether different decoding challenges are encountered by learners at different stages of L2 development.

To sum up this section, the above gaps in research prompted two questions: First, to what extent will Arabic L1 readers (Saudi foundation year ESL students in the present study) be able to perform successfully on measures of English decoding? Second, what is the nature of their L2 decoding realisations (outputs), experiences, and processes? Assessing the decoding abilities of Saudi foundation year students (CEFR: A1, A2, and B1) would allow us to know whether or not, and the extent to which, these students encounter difficulties with decoding their second language, English. Nonetheless, test scores offer no insights into the *nature* of any potential decoding difficulties, nor to the nature of any decoding processes that students engage in, which might contribute to the observed decoding realisations and scores. An investigation into readers' decoding realisations, experiences, and processes would provide greater depth to the obtained decoding proficiency scores, allowing for a comprehensive understanding of the decoding challenges among (Arabic L1) Saudi foundation year ESL students.

1.2. Research into the contribution of L2 decoding proficiency to L2 reading comprehension

The fact that prior studies have reported that L2 decoding skill is not easy to acquire, particularly in readers with dissimilar L1WS backgrounds, makes it more important to understand whether the decoding proficiency of these readers constrains the ultimate goal of their decoding: reading comprehension. To date, however, the contribution of L2 decoding proficiency to L2 reading comprehension remains unclear (Nassaji, 2014; Pasquarella, Gottardo, & Grant, 2012). The very few componential studies in the L2 reading context (Jeon, 2011) may, in the first place, explain this lack. However, even among existing componential studies of L2 reading comprehension, L2 decoding proficiency was a largely overlooked variable, with the contributions of L2 linguistic knowledge – operationalized as breadth of L2 vocabulary and grammar knowledge – and L1 reading comprehension ability receiving the bulk of researchers' attention (Zhang, 2012). It is thus unsurprising that, in her componential model of L2 reading comprehension, Bernhardt (2005) makes no explicit mention of L2 decoding ability. The model suggests that first language reading comprehension accounts for 14-20% of the variance in L2 reading comprehension, whereas L2 linguistic knowledge (vocabulary and grammar knowledge combined) accounts for 30-38% of the variance, leading Bernhardt to raise the question: what accounts for the remaining variance in L2 reading comprehension scores?

There remains, however, a handful of studies that investigated the contribution of L2 decoding proficiency to L2 reading comprehension. Nonetheless, these are limited by several issues, addressing which would have valuable implications not only for L2 reading theory, but also pedagogy. These issues are apparent in:

(i) Lack of key control variables (e.g., Koda, 1998), thus, making it difficult to: a) ascertain whether L2 decoding proficiency would make a distinct contribution to L2 reading comprehension, over and above the contributions of key predictors of L2 reading comprehension such as L2 vocabulary knowledge and L2 grammar knowledge (Nassaji, 2014); and b) evaluate the size of the decoding contribution, if any, with that of these previously established predictors of L2 reading comprehension.

(ii) Limitations in scope evident in the employment of readers from a single proficiency level (for example, intermediate readers *only* in Jiang et al. (2012)), even though, as will be argued later, the contribution of L2 decoding proficiency to L2 reading comprehension could potentially differ as a function of the proficiency level of the learner.

(iii) Limitations in assessment methods evident in the employment of single decoding measures, that is, real word *or* pseudoword decoding measures (for example, pseudoword *only* in Jeon, 2011). This not only restricts thorough assessment of decoding skill, but also makes it difficult to see whether the two decoding measures would make an independent contribution to L2 reading comprehension performance.

(iv) Focus on English as an additional language setting (i.e., language minority children in the English-speaking world), rather than on adult readers in second language settings (i.e., non-Enlish speaking countries). Unlike immigrant children who lack extensive decoding experience in their L1WS, adult L2 readers often have extensive L1 decoding experience that might influence their L2 decoding, and consequently, its contribution to L2 reading comprehension performance (Jeon and Yamashita, 2014).

(v) Lack of decoding speed indicators, though assessing speed can more accurately reflect the contribution of decoding proficiency to reading comprehension (Florit & Cain, 2011).

(vi) Lack of complementary qualitative investigations, though these may not only provide means of triangulation for quantitative correlational data, but also an indication of the mechanisms that underlie the contributory role of decoding to reading comprehension, as well as an indication of the factors that might account for any unexplained variance in the quantitative model tested.

(vii) Lack of heterogeneity among L1 backgrounds studied in the field as a whole, though extending the range of L1WSs would provide evidence for the cross-L1WS validity of the contribution of decoding to reading comprehension in an L2 (As will be reviewed later, Jeon and Yamashita (2014), showed how L1WS might influence the strength of the L2 decoding proficiency-L2 reading comprehension relation).

It is worth noting that the last limitation is of paramount relevance to the present study. Studying readers with L1 Arabic backgrounds would not simply provide evidence for the cross-L1WS validity of this contribution. It would also shed light on a link speculated on since Ryan and Meara's (1991) seminal work on Arab ESL readers' 'vowel blindness'. Vowel-related word reading difficulties have been presented as a key impediment to Arabic readers' L2 reading comprehension. However, to date, this link remains hypothetical as it is yet to be empirically tested. Hayes-Harb (2006, p.338) has indeed pointed out this limitation:

> To determine whether native Arabic speakers' exceptional treatment of vowel letters in English reading can explain their exceptional English reading comprehension difficulty, research is needed that directly investigates comprehension as it relates to issues of visual word processing.

1.3. The Saudi setting for ESL reading in Higher Education

The current education policy in Saudi Arabia stipulates that Arabic should be the medium of instruction in schools, and that English be taught as a school subject starting from grade 5. However, reflecting the global interest in English-Medium Instruction (EMI), English has become the language of instruction in most Saudi higher education institutions (Ebad, 2014). Not unsurprisingly, this shift has presented several challenges, described once as one that "takes a direct hit on [students'] knowledge, performance, and learning outcomes. Consequently, the results are poor grades, drop outs, and low achievers" (Ebad, 2014, p. 141).

To address these challenges, universities have established language centres which deliver compulsory English courses for foundation year students (ibid). At one higher education institution, the funder of the current project and which hosts the largest number of undergraduates in the country (Ministry of Education, 2017), all students are administered a placement test to assess their ESL proficiency before the foundation year commences (HEI, 2016). They are then enrolled in an ESL course designed to take them to the Intermediate Level (B1+ CEFR) by the end of the foundation year. Progressing to the B1+/intermediate level is a requirement to progress to the second year, where students are admitted to colleges (ibid).

Finally, a few remarks from my own teaching experience might be worth adding here. To every ESL practitioner in this site, it is unquestionable that students struggle with English reading comprehension. While work on improving students' vocabulary and grammar is underway, efforts towards improving their decoding ability are non-existent. This, however, is not surprising. Current pedagogical practice – in this site and potentially in many other L2 contexts – significantly hinders understanding of learners' decoding abilities and needs. Asking learners to read aloud is often discouraged on the view that it would provoke negative feelings in the L2 classroom (for example, anxiety). Further, even when reading aloud activities are incorporated in the classroom, it is often the skilled decoders who would be willing to participate in these activities. The issue here is that this could obscure, or worse, distort, our understanding of the decoding abilities, and needs, of these students. This is further compounded by the fact that decoding skill – unlike other language skills (for example, grammar, vocabulary) – is neglected in all forms of assessments undertaken as part of the ESL program. Thus, until researched, any difficulties these learners encounter with English decoding would remain invisible, and consequently unattended. Further, if their reading

comprehension difficulties are explained, at least in part, by their decoding performance, then it would be all the more important to work towards improving the decoding abilities of these students.

1.4. The aims of the present study:

The present study aims to:

(1) explore the L2 decoding performance of Saudi (Arabic L1) foundation year ESL students. Particularly, to explore how well they are able to perform on measures of L2 English decoding, and the nature of their decoding realisations (outputs), experiences, and processes.

(2) investigate the contribution of these learners' L2 decoding proficiency to their L2 reading comprehension, relative to the contributions of L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension.

To achieve the first aim, 216 Saudi foundation year ESL students took part in a real word and pseudoword reading aloud task aimed at assessing their decoding abilities. This data was then complemented by a qualitative investigation into readers' decoding outputs, experiences, and processes. Particularly, the decoding outputs of a subsample of 60 students were phonologically transcribed and analysed. The error analysis data was further complemented with semi-structured interviews aimed at investigating the decoding experiences of a subsample of 12 participants. Finally, a further subsample of 15 particpcats took part in a restrospective think aloud task which aimed at understanding the decoding processes that learners engaged in as they decoded English.

To achieve the second aim, the same 216 participants were administered measures of L2 reading comprehension ability, L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability. This allowed us to examine the contribution of L2 decoding proficiency to L2 reading comprehension, when variance attributed to these key predictors of L2 reading comprehension is taken into account, and also to compare the size of the decoding contribution with that of these previously-established predictors of L2 reading comprehension. The inclusion of learners at different levels of proficiency allowed us to see whether the contribution of L2 decoding proficiency varies as a function of the proficiency level of the learner, and if so, in what ways. Further, the inclusion of L2 real word and L2 pseudoword decoding measures allowed an investigation of the independent contributions (or lack thereof) of these two measures to L2 reading comprehension. The quantitative data obtained was also

complemented by a follow up qualitative investigation in which 18 participants were interviewd to reflect on their experiences of the issues that interfere with their reading comprehesnion performance, with a view to see whether decoding-related issues would emerge as a contributory factor, and if so, why.

1.5. The structure of the thesis

The present thesis contains seven chapters. The structure of each is described below:

1) Introduction

To set the overall argument of the thesis, this first chapter has briefly introduced previous research into second language decoding as well as research into the contribution of L2 decoding proficiency to L2 reading comprehension ability. It has highlighted limitations present in these two lines of research and thus provided the rationale for the focus of the present study. Then, following a description of the context of the current study, the study's aims have been outlined.

2) Literature review

The second chapter is divided into two main parts. The first part defines what reading comprehension means and provides a review of current reading comprehension models. It details the components crucial for successful reading comprehension, with particular focus on the component of decoding ability and the mechanisms that might underpin its contribution to reading comprehension. It then provides a detailed discussion of gaps in previous research into the contribution of L2 decoding proficiency to L2 reading comprehension performance.

The second part of the chapter presents how L2 decoding ability has been understood. It reviews various L2 decoding research strands including quasi experimental, descriptive, and qualitative studies, and highlights limitations present in this line of research. The section concludes by highlighting gaps in prior research into the decoding challenges encountered by second language learners. Finally, it identifies the unique position of the current study in relation to previous studies and presents the research questions.

3) Methodology

The third chapter explains the research methodology of the current study. It first describes the study's design. Then, it outlines the sampling procedure (including a description of the sampling frame, sampling size, and sampling strategy). The study's overall procedure is then presented. This is followed by a full description of the quantitative and qualitative data collection methods. The chapter then proceeds to present the data analytical techniques adopted

to answer each of the research questions. Finally, following a discussion of the piloting phase, a discussion of ethical considerations is presented.

4) Results I

The fourth chapter presents the findings of the first part of the study and is divided into four sections. The first section presents the findings of the analysis into the real word, pseudoword, and overall decoding abilities of Saudi foundation year ESL students (CEFR=A1, A2, B1). The second section presents a detailed analysis of the decoding realisations (outputs) of these learners. Particularly, it presents an overview of the main and sub-categories of errors observed across the study participants, and whether these findings differ in learners at different levels of proficiency. After that, the third section provides the findings of the analysis into learners' experiences with English decoding. The fourth and final section presents an overview of the decoding processes observed across the study participants.

5) Results II

The fifth chapter presents the findings of the second part of the study. Following initial quantitative data screening and some preliminary analyses, it reports the relationships between overall L2 decoding proficiency, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension ability. Secondly, it presents the findings of a regression analysis into the contribution of overall L2 decoding proficiency to L2 reading comprehension before and after controlling for the contributions of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability. Thirdly, it presents the findings of an analysis into the contribution of overall L2 decoding proficiency as a function of the proficiency level of the L2 learner. Fourthly, it presents the findings of an analysis into the independent contributions (or lack thereof) of the L2 real word decoding and L2 pseudoword decoding measures to L2 reading comprehension performance. The chapter concludes with some qualitative insights into the contribution of L2 decoding proficiency to L2 reading comprehension derived from semi-structured interviews with the study participants.

6) Discussion

The sixth chapter discusses the findings of each of the research questions in the light of existing theory and literature. Then, it discusses the implications of these findings for theories of reading and for second language pedagogy. The limitations of the present study are also discussed in this chapter.

7) Conclusions

The seventh and final chapter provides a brief summary of the present study as well as some recommendations for future avenues of research.

2. Literature Review

This chapter is divided into three sections. The first section is concerned with the contribution of L2 decoding proficiency to L2 reading comprehension. It opens with an introduction to what reading comprehension means and provides a descriptive and critical account of current reading comprehension models, including "process" and "componential" models, with a particular emphasis on the latter due to their relevance to the present work. It then lists the components of L2 reading comprehension, as established in previous research, with special focus on the component of L2 decoding proficiency. Finally, the first section concludes with a discussion of gaps in prior research into the contribution of L2 decoding proficiency to L2 reading comprehension.

The second section deals with acquisition of second language decoding skill, and more particularly the challenges that second language readers encounter along the path of L2 decoding acquisition. It begins by introducing the concept of a writing system (WS) and how it shapes the nature of decoding. It then presents Koda's (2008) orthographic Distance Hypothesis along with its empirical underpinnings. The second section concludes by highlighting gaps in prior research into the decoding challenges encountered by second language learners.

The third and final section presents the aims of the present study and the research questions it set out to answer.

2.1.Models of reading comprhension:

Reading comprehension has been described as the product of interaction between the writer of the text and the reader (Kingston, 1961). Put simply, it reflects the reader's "ability to draw meaning from the printed page and interpret this information appropriately" (Grabe & Stoller, 2011, p. 9). On the surface, this definition may give the impression of reading comprehnesion being a simple activity; however, the scientific consensus suggests that it is far from one. And, over the past few decades, various theoritical models have been proposed to explain this largely complex activity. These models generally fall into two categories: "Process Models of reading comprehension" and "componential models of reading comprehension" (Kremmel, Brunfaut, & Alderson, 2015; Shiotsu, 2010). The former are mainly concerned with explaining the exact process of reading comprehesnion, whereas the latter are concerned with explaining individual differences in reading comprehension outcomes (ibid). Both classes of model are reviewed in the following section.

2.1.1. Process models of reading comprehension

Process models, also termed "metaphorical models" (Grabe & Stoller, 2011, p. 31) make assumptions about *how* the process of reading comprehension is carried out (Grabe, 2009). These include the so called "bottom up", "top-down", and "interactive" models of reading. "Bottom-up models" depict reading as an activity which comprises a low-level stage (including processes such as word decoding) and a higher-level stage (including processes such as inferencing, prediction, etc.), with readers proceeding through these stages in a sequential bottom-up fashion (Gough, 1972; LaBerge & Samuels, 1974). In other words, lower-level processes operate almost independently from higher-level processes (i.e., with little to no interference from higher-level operations; for example, a reader's expectation about the text is not assumed to influence their word reading). A strict bottom-up framework is no longer accepted as valid as it failed to account for several findings in the reading literature, such as the effects of the sentential context on word reading (Stanovich, 1980, 1986, 2000).

"Top-down models", as their name suggests, view reading in a way that contrasts with that of bottom-up models. According to these models, higher-level processing operations direct and influence information obtained from lower-level operations. Thus, reading is viewed as a "psycholinguistic guessing game" in which the reader formulates certain expectations about the text they are reading and engages in continual hypothesis testing, as they proceed through that text, to verify these expectations (Goodman, 1967; Smith, 1978). This view of reading was also critiqued on logical and empirical grounds. To illustrate, forming hypotheses about upcoming words by relying on syntactic and semantic analyses is likely to require more processing time than the milliseconds often needed to convert graphemes into phonemes (Grabe, 2009). Furthermore, the assumptions of the model have not been borne out by empirical findings, as shown in research of readers' eye movements as well as research on context effects on word reading. That is, there is evidence that even skilled readers fixate on nearly every word during reading, though they do so quickly and efficiently (Rayner & Pollatsek, 1989). There is also evidence that skilled readers do not tend to make use of expectations or sentential context to aid word reading (Stanovich, 2000).

A third class of process models, "Interactive Models of Reading", served as the middle ground between these two opposing views. In these models, the reader is assumed to collect information about the text by utilizing lower-level and higher-level processing operations *simultaneously* (as opposed to a sequential/stage-like bottom-up or top-down manner). This moderate stance appears to be more acceptable in the literature than extreme bottom-up and top-down views. However, it was problematic in one respect: the model assumes that the *nature*

of that interactive engagement (of lower-level and higher-level operations) is similar for all readers (Stanovich, 1986). However, research has consistently shown that readers differ with respect to the *degree* of involvement of these operations across several reading-related tasks (Stanovich, 1980, 1986, 2000). For example, research has shown that skilled readers make less use of the sentential context (e.g., inferencing, a higher level process) to facilitate word reading (a lower-level process), wherease the opposite was observed for poor readers (ibid).

Against this backdrop, Stanovich (1980, 1986, 2000) proposed a slightly modified version of the interactive model. His "Interactive Compensatory Model" assumes that higher-level and lower-level operations operate in a compensatory (rather than random) manner. In other words, when one operation is less efficient, other operations will compensate for this inefficiency to ensure that comprehension is not interrupted. For example, if a reader encounters difficulties reading words (a lower-level operation), they will use contextual information (a higher-level operation) to faciltate the decoding process, a finding supported by previous research (ibid).

To sum up this section, process models (including bottom-up, top-down, interactive, and interactive compensatory models) are generally concerned with explaining how the process of reading comprehension is carried out, including the time course of lower-level and higher-level operations. Strict bottom-up and top-down views are no longer viewed as valid, whereas an interactive view, and perhaps particularly that with a comensatory component, continues to be seen as valid in the literature.

Process models were useful as they postulated fundamental processing ideas about reading comprehension (ibid). However, these models tell us little, if anything, about individual differences in reading comprehesion *outcomes* (Shiotsu, 2010). In the next section, a discussion of another line of reading comrpehension models is presented, "Componential Models of reading comprehension", followed by a review of two prominent componential models. The components proposed to be influential to L2 reading comprehension are then discussed along with relevant empirical evidence. When discussing the latter, greater emphasis (and space) is devoted to the contribution of decoding ability, the component of interest in the present study. Furthermore, because the present study is concerned with second (rather than first) language reading, only components proposed to be influential in the second language context are reviewed in this section.

2.1.2. Componential approaches of reading comprehension

Componential models are more concerned with identifying the possible explanatory variables (components) in reading comprehension, as opposed to explaining how these

components operate in the process (Kremmel, Brunfaut, & Alderson, 2015; Shiotsu & Weir, 2007). Reading comprehension is viewed as a complex operation that involves many knowledge components (such as decoding knowledge, vocabulary knowledge, grammar knowledge, etc.), any of which can be a source of individual differences in reading comprehension outcomes. To put it simply, this body of research is concerned with examining whether, and the extent to which, various reading components (mostly operationalised using relevant language-related tests) can predict variability in reading comprehension ability (ibid). As such, componential approaches were viewed as valuable for theory and pedagogy. By "test[ing] and compar[ing] the relative importance of theoretically hypothesized components, [they] offer insights into the construct of reading ability, facilitate understanding of individual differences, and ultimately provide more balanced implications for language pedagogy and assessment" (Yamashita & Shiotsu, 2015, p. 1).

In what follows, a review of two prominent componential models is presented. The first is the "Simple View of Reading", proposed to explain variations in *first* language reading comprehension (though see below), and the second is Bernhardt's model of reading, proposed to explain variations in *second* language reading comprehension performance.

2.1.2.1.Review of existing componential models:

2.1.2.1.1. The Simple View of Reading

Over three decades ago, Gough and Tunmer (1986) proposed "the Simple View of Reading" in an attempt to explain individual differences in reading comprehension performance. The central premise of the SVR is that variations in reading comprehension ability can be primarily explained by variations in two subskills: decoding (typically assessed using pseudoword and/or real word naming tasks) and linguistic comprehension (typically assessed using a listening comprehension task). The linguistic comprehension component is assumed to reflect vocabulary knowledge, syntactic knowledge, and general sentence and paragraph comprehension abilities, for example, inferencing skills (Hoover & Gough, 1990; Kirby & Savage, 2008). The Simple View also assumes that neither of these two components is sufficient by itself, as both are necessary for comprehension to occur. This assumption was expressed in the formula: Reading comprehension = Decoding x Linguistic comprehension, or R= D x C. Thus, if a learner has zero decoding skill, their reading comprehension will equal zero even if they exhibit average-level listening comprehension skill. Similarly, if a reader has

zero linguistic comprehension, their reading comprehension will equal zero even if they have adequate levels of decoding skill. Hoover and Gough (1990) also noted that the contribution of these two components will change during the course of development. In young readers, decoding will contribute more significantly to reading comprehension than oral linguistic comprehension, whereas in older readers, and as decoding skill reaches ceiling level and reading materials become more complex, variability in linguistic comprehension will account for most of the variation in reading comprehension skill.

The power of the SVR that most reading researchers acknowledge is its simplicity (though see below). It provides a parsimonious explanation for individual differences in a seemingly complex skill (Chen & Vellutino, 1997; Grabe, 2009; Nation, 2019). Furthermore, it provides researchers with testable hypotheses which can inform theoretical understanding and, subsequently, pedagogical practices (Chen & Vellutino, 1997). Finally, the basic essence of the theory is well supported by empirical evidence. In other words, several studies have shown that variations in decoding and linguistic comprehension were strongly correlated with variations in reading comprehension skill (Nation, 2019). However, it should be noted that most of this evidence comes from first language reading studies (Kirby & Savage, 2008). In the few studies involving second language learners, participants were often immigrant children already residing in the second language speaking country (e.g., Verhoeven & van Leeuwe, 2012).

However, despite the above strengths, the SVR is not without problems. Firstly, while empirical support exists for the predictive power of decoding and linguistic comprehension, controversy exists as to how the two skills combine to predict reading comprehension. To illustrate, Gough and Tunmer (1986) and Hoover and Gough (1990) proposed that a multiplicative equation (i.e., R= D x C) can better capture variations in reading comprehension than an additive equation (i.e. R= D + C). However, several studies either found that: (a) the product term (DXC) did not explain any unique variance in reading comprehension after the additive term (D+C) was included in the analysis (i.e., after decoding and listening comprehension have been controlled for), and/or that (b) both the product and additive term predicted reading comprehension to a similar extent (Chen & Vellutino, 1997; Georgiou, Das, & Hayward, 2008; Joshi & Aaron, 2000; Savage & Wolforth, 2007; Silverman, Speece, Harring, & Ritchey, 2013). Georgiou, Das, and Hayward (2008) argued that these results suggest that decoding and listening comprehension are sufficient, but *not necessary* for reading

comprehension to occur, and that one of the two can be bypassed in reading comprehension. However, a more conceptually plausible explanation for these findings was found in Silverman, Speece, Harring, and Ritchey (2013) and Kirby and Savage (2008). They argued that for normal readers, where decoding or listening comprehension are unlikely to be zero, the additive model would be as appropriate as the multiplicative model, but that for severely impaired readers whose decoding or linguistic comprehension skill could approach zero, the multiplicative model would be more appropriate.

Secondly, at least as far as the developmental pattern of the contributions of decoding and linguistic comprehension is concerned, the SVR does not seem to be applicable to readers of other alphabetic writing systems. For example, Florit and Cain's (2011) meta-analysis found that, throughout the course of literacy development, the relative contributions of decoding and linguistic comprehension were different for readers of different writing systems. Particularly, they found that the contribution of linguistic comprehension appeared to be more critical to reading comprehension than that of decoding accuracy in beginning readers of transparent orthographies (an opposite trend was found for beginning readers of English, i.e., decoding was more detrimental to reading comprehension in the early stages of literacy development) (ibid).

Thirdly, the linguistic/listening comprehension component of the SVR is problematic for two reasons: (1) it has a low explanatory value, and consequently, limited impact on theoretical understanding and pedagogical practice. For example, it makes it difficult to understand, and compare, the individual contributions of the hypothesized constituents of "linguistic comprehension" (for example, vocabulary knowledge, grammar knowledge, inferencing abilities, etc.); (2) on purely logical grounds, it can be argued that it is somewhat absurd to conceptualize listening comprehension as a component of reading comprehension. As Gernsbacher, Varner, and Faust (1990) noted, the "high correlations between listening and reading suggest a General Comprehension Skill factor that transcends modality". Thus, since the two constitute different modalities of the same phenomenon (language comprehension), assigning one as a predictor of another may not be meaningful (i.e., assigning language comprehension as a predictor of language comprehension). Put differently, this is not different from saying that language comprehension is related to language comprehension (which is somewhat a form of circular reasoning that does not conform well with a basic law of causality: a cause cannot cause itself). This is not to mention that this redundancy could potentially impact

statistical findings by resulting in artificially inflated correlation coefficients. A more theoretically, pedagogically, and logically meaningful approach would be to unpack the listening comprehension component into sub-components and assess the individual contributions of these to reading comprehension.

Finally, though the SVR has been around for 30 years, it is interesting to note that confusion seems to exist as to whether it is a model of first or second language reading (or both). This is possibly because the SVR was first "verbally" described in Gough and Tunmer (1986), and then empirically tested later using secondary data which comprised US-based "bilingual' Spanish-English children (Hoover and Gough, 1990). Hoover and Gough (1990) did not provide sufficient information about the L1 or L2 proficiency levels of their participants but did note that that there was variation in the degree of bilingualism among the participants (i.e., some were fluent users of English but not Spanish, others were fluent users of Spanish but not English, and some were fluent users of both English and Spanish). It is thus unsurprising to see this confusion in several descriptive accounts of the model. For example, Grabe (2009) and Kirby and Savage (2008) appear to view it as a model of first language reading (as evidenced by their questioning of the validity of the model in second language settings). Others (Silverman, Speece, Harring, & Ritchey, 2013) seemed to view it as a model of second language reading (as explicitly stated in their description "Hoover and Gough's (1990) study of second language learners"). Others still (e.g., Chen, 1997, and Joshi (2000)) view it as a model for "bilingual" readers.

In what follows, a componential model that was specifically developed to explain variability in second language reading is reviewed.

2.1.2.1.2. Bernhardt's componential model of L2 reading comprehension

In the L2 reading context, as Grabe (2009) noted, one main model exists, and this is the model put forward by Bernhardt (2000; 2005). This model was the product of intense debate among L2 reading researchers (McNeil, 2012). Following Alderson's (1984) question of whether L2 reading is a reading problem or a language problem, several studies investigated, and compared, the impact of first language reading comprehension, and L2 linguistic knowledge (mainly L2 grammar and vocabulary knowledge), on L2 reading comprehension performance (e.g., Bernhardt & Kamil, 1995). Bernhardt's (2000; 2005) review of that literature found that L1 reading explained around 20% of the variance in L2 reading

comprehension whereas L2 linguistic knowledge explained around 30% of the variance in L2 reading comprehension (though see sections 2.1.2.2.2, 2.1.2.2.2., and 2.1.2.2.3). In light of this, Bernhardt (2000; 2005) noted that L1 reading models (which naturally did not include an L1 reading component) may not be readily applicable to the L2 reading context. Thus, an L2 model which acknowledges the contribution of L1 reading ability, alongside that of L2 linguistic knowledge, was needed. The central premise of her model was that variations in second language reading ability are mostly explained by variations in L1 reading ability (accounting for ~20% of the variance), and L2 grammar and vocabulary knowledge (accounting for ~30% of the variance), with the remaining 50% of the variance left "unexplained" (Bernhardt, 2000, 2005). Furthermore, the model also includes a compensatory component to suggest that deficiencies in any reading component/subskill can be overcome by relying on other components/subskills (Bernhardt, 2005).

Bernhardt (2000) noted that her model has several strengths: First, it acknowledges the contribution of L1 reading ability when the L2 field was primarily reliant on L1-based understandings of L2 reading; Second, Bernhardt (2000) argues that the model breaks down the complex skill of L2 reading comprehension into components, which enhances theoretical understanding of L2 reading; Third, it "promotes the consideration of unexplained variance in individual performance and after considerable time in instruction" (p.804) (though it is worth noting here that Bernhardt (2005) proposed that the "unexplained variance" component could potentially reflect variations in "other" domains including affect, background knowledge, and comprehension strategies).

Nonetheless, the model has been critiqued for its vague identification of component skills (Grabe, 2009). For example, Bernhardt (2000; 2005) simply noted that the linguistic knowledge component refers to "vocabulary" and "grammar" knowledge providing no information about what these two constructs exactly mean. Given that vocabulary knowledge is multifaceted (for example, involving word meaning knowledge, word form knowledge, collocational knowledge, etc.), it remains unclear which aspects of word knowledge underpin the vocabulary component in the model (though it is highly likely that the component reflected word meaning knowledge as this is how the studies reviewed operationalized vocabulary knowledge). Furthermore, in Bernhardt (2005), the linguistic knowledge component seemed to additionally reflect "cognates and L1/L2 linguistic distance" (p.140), yet, again it is unclear what "cognates" and "linguistic distance" exactly mean. Confusion also surrounds the "L1 reading" component (for example, whether it reflects L1 reading comprehension ability or L1

reading accuracy or both). Additionally, the "unexplained variance" leaves much to be known about variability in L2 reading comprehension (McNeil, 2012).

Furthermore, while the empirical literature generally supports the central assumption of the model (i.e., that linguistic knowledge and L1 reading ability explain variations in L2 reading comprehension ability), the exact amount of variability explained by these components varied considerably from what the model assumed. For example, in Yamashita and Shiotsu (2015), linguistic knowledge (defined as grammar and vocabulary knowledge) explained around 74% of the variance in L2 reading comprehension. Similarly, in Lee and Schallert (1997), nearly 30% of the variability in L2 reading comprehension was attributed to L1 reading ability alone.

Finally, and most importantly, the model did not include a decoding component at all despite the potential role of L2 decoding proficiency in L2 reading comprehension (see section 2.1.2.2.6). This is, however, unsurprising given that (1) L2 decoding skill received little attention in the SLA field in general (Koda, 2005); and (2) the field's heavy emphasis on top-down views of reading which undermined the role of lower-level processes (Nassaji, 2014).

2.1.2.2.Components (predictors) of L2 reading comprehension ability

The L2 reading literature highlighted several predictors of L2 reading comprehension outcomes. In what follows, these predictors are presented along with relevant theoretical and empirical evidence. Due to space limitations, selection of the components to be reviewed was guided by the findings of Jeon and Yamashita's (2014) meta-analysis on L2 reading comprehension correlates. This meta-analysis examined the strength of correlation between L2 reading comprehension performance and 10 predictors of L2 reading comprehension: L2 vocabulary knowledge, L2 grammar knowledge, L2 decoding proficiency, L1 reading comprehension, L2 listening comprehension, L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, working memory, and metacognition. Because the last five components were classified as "low-evidence" correlates in Jeon and Yamashita's (2014) work, *and* because they also constitute subcomponents of other reading components (see section 2.1.2.2.5), these five components are merged and discussed in an "other components" section (section 2.1.2.2.5).

Finally, for organizational purposes, the decoding component is presented last to allow for thorough discussion of relevant theoretical and empirical evidence, along with a discussion of the limitations in prior research into the contribution of L2 decoding proficiency to L2 reading comprehension performance.

2.1.2.2.1. L2 Vocabulary knowledge (L2VK)

As noted previously, vocabulary knowledge is multifaceted, but central to it is knowledge of word meaning (Koda, 2005). Furthermore, in reading research where reading comprehension is the dependent variable, vocabulary knowledge is often defined as the breadth of word knowledge, i.e., how many words a reader knows the meaning of (Jeon, 2011; Perfetti, 2010).

From a theoretical standpoint, it is intuitive that understanding individual word meanings is pivotal in overall passage meaning construction (Koda, 2005). Empirically, several studies reported high correlations between vocabulary knowledge and reading comprehension (Nassaji, 2003; Qian & Schedl, 2004; Shiotsu, 2010; Zhang, 2012). In Jeon and Yamashita's metanalysis, vocabulary emerged as the second strongest correlate of L2 reading comprehension with an overall mean correlation of r=.79 (following L2 grammar knowledge, though there was a large overlap in their 95% confidence intervals, leading the researchers to argue that both skills are equally important for L2 reading comprehension). Apart from correlational/prediction studies, direct evidence for the contribution of L2 vocabulary to reading comprehension was also found in intervention studies (e.g., Lesaux, Kieffer, Faller, & Kelley, 2010).

2.1.2.2.2. L2 Grammar knowledge (L2GK)

In reading research, grammar knowledge was most commonly defined as morphosyntactic knowledge (Jeon & Yamashita, 2014, p. 165). That is, "knowledge about a variety of morphosyntactic properties, i.e. tense, aspect, word order, subject–verb agreement, and articles, etc." (Jeon & Yamashita, 2014, p. 165).

It has been argued that poor grammar knowledge could exert an additional burden on L2 readers. This is because grammar knowledge allows readers to analyze and integrate syntactic information, thus providing ongoing instructions for the construction of passage comprehension (ibid; Grabe, 2009). Empirically, associations between grammar knowledge and reading comprehension are well established in the L2 reading literature (e.g., Shiotsu, 2010; Zhang, 2012). In Jeon and Yamashita's metanalysis, grammar knowledge emerged as the strongest correlate of L2 reading comprehension with an overall mean correlation of r=.85.

2.1.2.2.3. L1 reading comprehension (L1RC)

Associations between L2 reading comprehension ability and L1 reading comprehension ability have been frequently reported in the literature. Traditional theoretical understanding for the contributing role of L1 reading comprehension is derived from the Common Underlying Proficiency hypothesis (Cummins, 1979, 1991) (also referred to as the "Interdependence hypothesis") and the assumption that general cognitive and comprehension skills (for example, ability to infer, predict, problem-solving skills, etc.) are not language specific and, thus, can be transferred – or "accessed", see Walter (2007) – across languages. However, it is interesting to note here that evidence exists for the role of the L1 writing system (and its relationship to that of the L2) in moderating the contribution of L1 reading comprehension to L2 reading comprehension. For example, in the meta-analysis by Jeon and Yamashita (2014), the overall mean correlation between L1 reading comprehension and L2 reading comprehension ability was r=.50. However, a follow up moderation analysis found that the relationship was stronger when the L1 and L2 share the same writing system (r=.60) than when the L1 and L2 writing systems differ (r=.36). This might suggest that the underlying explanation for the contribution of L1 reading comprehension goes beyond that of shared comprehension and cognitive skills to also represent variability in L1 word reading ability. In fact, some reading researchers seem to endorse this view. For example, McNeil (2012) mentioned that the L1 reading component could also reflect knowledge of letter-to-sound mapping (which reflects word reading ability). Finally, it is difficult to discuss the contribution of L1 reading without giving mention to the Language Threshold Hypothesis. This hypothesis suggests that L2 readers would need to have adequate levels of L2 knowledge before they are able to benefit from their L1 reading comprehension skills (Grabe & Stoller, 2011).

2.1.2.2.4. L2 listening comprehension (L2LC)

Some researchers considered L2 listening comprehension a component of L2 reading comprehension ability. This view is derived from the Simple View of Reading framework which characterises reading as the product of decoding ability and listening comprehension ability (Yamashita & Shiotsu, 2015). As mentioned previously, the rationale behind including listening comprehension as a component of reading is that the former can be an index of overall linguistic comprehension ability, thus representing vocabulary knowledge, grammar knowledge, and other general comprehension skills (for example, inferencing, predicting) (Kirby & Savage, 2008).

Several L2 studies reported positive associations between L2 listening comprehension and L2 reading comprehension (e.g., Jeon, 2011; Verhoeven & van Leeuwe, 2012). Most of these studies involved language minority children (e.g., Verhoeven & van Leeuwe, 2012) as these readers acquire L2 listening skills prior to literacy skills, but similar findings were also obtained from adolescent learners residing in their L1 countries (e.g., Jeon, 2011) (Yamashita & Shiotsu, 2015). In Jeon and Yamashita's (2014) meta-analysis, listening comprehension ability emerged as the third strongest correlate of L2 reading comprehension with an overall mean correlation of r=.77. Despite this large correlation, however, and as mentioned previously, the researcher has reservations about conceptualizing listening comprehension as a component of reading comprehension (see section 2.1.2.1.1), and it is not clear that much can be gained, both on a theoretical and pedagogical level, from continuing to view listening comprehension as a predictor of reading comprehension.

2.1.2.2.5. Other components

Apart from the predictors discussed above (and decoding, to be discussed next), other components have been posited to be predictive of L2 reading comprehension. These include phonological awareness, orthographic awareness, morphological awareness, working memory, and metacognitive awareness. As noted earlier, these variables were merged in this section for two reasons (apart from space limitations): Firstly, evidence for their predictive power is modest (see Jeon & Yamashita, 2014); Secondly, and most importantly, these components have been argued to be constituents of one (or more) of the components presented above (or decoding, to be discussed next). In other words, efficiency in one of the components presented above (or decoding, to be discussed next) could inherently reflect efficiency in one or more of the components in this "other" section.

For example, phonological awareness, orthographic knowledge, and morphological knowledge were considered subcomponents of decoding proficiency (see Adams, 1990; Kirby & Savage, 2008; Nassaji, 2014; Nation, 2019; Silverman, Speece, Harring, & Ritchey, 2013; Yamashita & Shiotsu, 2015). In other words, proficiency in decoding would intuitively reflect variability in learners' morphological, orthographic, and phonological awareness. In fact, there is evidence that phonological awareness does not account for any unique variance in reading comprehension when decoding proficiency is controlled for (Georgiou, Das, & Hayward, 2008).

Furthermore, morphological awareness has been found to be a significant predictor for vocabulary knowledge, and in research where reading comprehension was the criterion of interest, the contribution of morphological awareness was found to be fully mediated by vocabulary knowledge (Zhang and Koda, 2012) (though see Kieffer and Lesaux (2012), where morphological awareness continued to make a statistically significant, yet very small, contribution to reading comprehension even after controlling for vocabulary knowledge).

Finally, it has been argued that working memory, metacognitive awareness (and L1 reading comprehension ability), are all representative of language-general cognitive and metacognitive skills (Jeon & Yamashita, 2014). Thus, it can be argued that variability in working memory and metacognitive awareness could be largely (though possibly not fully) captured in the L1 reading comprehension component (see Grabe and Stoller, 2011).

2.1.2.2.6. Decoding proficiency

As one part of the present study is concerned with investigating the contribution of L2 decoding to L2 reading comprehension, more space is dedicated to the decoding component here. The section begins by defining decoding proficiency. It then discusses the theoretical underpinning of its role in reading comprehension and follows that by a review of relevant empirical evidence. The section concludes by highlighting gaps in previous research into the contribution of L2 decoding to L2 reading comprehension performance.

2.1.2.2.6.1. Definition

Decoding proficiency has been defined (and operationalized) in many ways in the literature. For example, some researchers equated it with accuracy and/or speed of real word reading (e.g., Dreyer & Katz, 1992); some with accuracy and/or speed of pseudoword reading (e.g., Jeon, 2011; Koda, 1998); and some with accuracy and/or speed of both real word and pseudoword reading (e.g., Jiang, Sawaki, & Sabatini, 2012; Kirby & Savage, 2008).

The term has also been interchangeably used with word recognition or word identification. For example, Hoover and Gough (1990) define decoding as "efficient word recognition" (p. 130). Nonetheless, in the latter case, researchers often refer to learners' ability to read words correctly regardless of whether that reading was associated with accurate meaning access. For example, Cunningham (1993, p. 32) defines "word identification" as "the process of making the print-to-sound link for each word encountered in print, regardless of whether this is done orally, subvocally, or neurologically and whether or not meaning is also accessed".

In the present study, decoding proficiency refers to the ability to transfer written codes, e.g., <cat>, to phonological codes, e.g., /kæt/ accurately and quickly. As will be discussed, phonological codes can be generated by breaking the whole-word graphical code into units, computing the sound of each unit until the whole-word phonological code is generated, or, by accessing the whole-word phonological code immediately from long-term memory (Coltheart, Curtis, Atkins, & Haller, 1993). Whenever decoding is used in this research, it denotes the generation of phonological codes broadly regardless of the method of that generation.

In the case of reading comprehension, decoding is assumed to operate subvocally and/or neurologically (i.e., as inner speech) (Cunningham, 1993; Perfetti, Bell, & Delaney, 1988). In what follows, theoretical support for the role of decoding in reading comprehension is discussed.

2.1.2.2.6.2.Theoretical support

Three explanations currently exist for the contribution of decoding proficiency to real-time reading comprehension. **The first explanation** relates to the support given by inner speech in working memory. To begin with, it is now established that phonological memory has a larger capacity than visual memory, and that phonological information (i.e., word sound) is far less susceptible to quick decay than visual information (i.e., word shape) (Coltheart, 2006). As such, phonological information could enhance information processing when they are referenced for use in higher level operations, as expressed by Perfetti, Zhang, and Berent (1992, p. 229)

> Semantic information abstracted from word forms is not enough. The functional difference between dog and canine is their distinct phonological forms, not their meanings. It is clear that readers in fact generally retain the exact wording for a sentence or at least a clause, just as listeners do. This verbatim retention allows a representation over which to parse sentences, repair parsing failures, integrate propositions and make inferences.

It seems plausible then to assume that if these phonological codes were inaccurate, referencing them would do little to facilitate the processing of text, and might potentially disturb the flow of comprehension. As Walter (2008) exemplified, an L2 reader decoding <white> as \wi:t\ would risk confusing <white> with <wheat>, as the phonological codes of the two concepts would be indistinguishable in their phonological loop.

Furthermore, since phonological codes are useful in holding words, and word order, in working memory (as they are being processed) (Hanson, Goodell, & Perfetti, 1991), readers

avoiding decoding words, for difficulty reasons, might risk disturbing this word assembly process. Additionally, phonological codes are useful for recreating prosodic aspects (such as phrase or clause boundaries) which could aid comprehension (Pollatsek, 2015). Thus, again, readers who choose to bypass difficult-to-decode words might risk disturbing prosodic building, which in turn could impact comprehension.

The second explanation for the role of decoding in reading comprehension concerns an activity that occurs a little earlier in the reading comprehension process. Specifically, this explanation stems from the role of phonology in facilitating semantic access, a role that has been implicated in most word-reading theories. For example, in the prominent dual-route model, decoding is considered a key component of the meaning access process (Coltheart, Curtis, Atkins, & Haller, 1993). In one route, the "indirect" route, meaning access is thought to occur following computing the phonological form of a graphic shape by breaking it into units and applying relevant symbol-sound knowledge (and hence, is assumed as the 'less automatic' route). This route, and decoding skill, is useful in situations where the phonological representation of a word exists in long-term memory but is yet to be matched with its graphical code, as in the case of a reader matching a newly encountered written code with a pre-held phonological representation. It is also useful in situations where a reader is faced with an unfamiliar word, for which s/he has no pre-existing phonological representation, as in the case of new words/pseudowords.

In the other, "direct" route, the graphic shape of the whole word is mapped onto the whole-word phonological form that is pre-existing in long-term memory (and hence, is assumed as the 'more automatic' route). Though it was originally assumed that the phonological code does not facilitate/disrupt meaning access following the "direct" route, research has consistently found otherwise. The widely-cited study in relation to this argument is Van Orden's (1987). In this study, skilled readers (previously assumed to access meaning without phonological mediation), were found to be disrupted (exhibiting errors and slow speed) when presented with pseudohomophones such as *'rows'* and asked whether it formed a part of a category exemplar *'flower'*. The finding that readers' accuracy and speed on a semantic judgment task was hindered by phonological interference was taken as evidence for the role of decoding in facilitating meaning access even when the "direct" route was employed. Similar evidence was also obtained from other research paradigms. For example, research on deaf readers and "tongue twisters" provided a test for the role of phonological decoding in reading. In Hanson, Goodell, and Perfetti's (1991) study, a semantic acceptability task was used to examine whether deaf (and hearing) college-level students use phonological decoding when reading.

Their results showed that both the hearing and deaf groups made more errors on semantic acceptability judgements when reading sentences involving tongue twisters than when reading control sentences. The authors argued that the tongue-twister effects provide evidence for the role of phonological decoding in reading. In general, these findings and several others suggested that graphic cues (irrespective of their familiarity) tend to build phonological activation prior to meaning access, and that phonological decoding can, therefore, mediate/disrupt meaning access (Grabe, 2009). This is why it was later suggested that it would be more accurate to conceive of the direct route as one which links the whole-word graphic code to the whole-word phonological code held in memory *and* that this whole-word phonological code will influence the meaning access process (Olson, Forsberg, Wise, & Rack, 1994). This is also why the pre-/post- lexical distinction, and by extension, the original conception of the dual-route model, has been questioned by many researchers, as Perfetti, Zhang, and Berent (1992, pp. 228-230) put it:

> Word identification research has been preoccupied with question #1 [when does phonological activation occur?]. Is phonological activation pre-lexical or post-lexical? This question is essentially whether phonology mediates the identification of a word (pre-lexical) or follows its recognition (post-lexical)…Our answer comes to this: Phonological activation always begins "pre-lexically". This activation always plays some part in identifying the word, provided the writing system allows it to do so... If phonological activation begins pre-lexically and is fully general across word types, then any model that implies strategic choice of routes is wrong... If what dual routes amount to is generalized phonological activation along both "routes," as most recent versions of dual route theory seem to allow, then they are compatible with our proposal.

The role of phonological decoding in semantic access is also acknowledged in what is known as a connectionist model of word processing. This model consist of networks of processing units: orthographic (e.g., letter/letter string), phonological (e.g., phonemes) and semantics, along with a set of hidden units that connect these elements (Seidenberg, 2007; Seidenberg & McClelland, 1989). The model learns to perform tasks such as decoding through repeated exposure to a range of words/spelling-sound mappings. Every time the model decodes a word, the decoding output generated is matched against the correct phonological form and, where inaccurate, that output is revised accordingly. The more frequently a particular grapheme-phoneme mapping is experienced, the stronger the links holding the orthographic and phonological units together. Unlike the original conception of the dual-route model which assumed that the two routes to semantic access are independent, connectionist models use information from both routes to read words, though to differing degrees (Seidenberg, 2007). In other words, early in training, semantic access is achieved following the activation of connections between orthographic units and phonological units. Following repeated exposure to words, however, the connections between orthography and semantics strengthen. However,

even in this case, decoding words continues to involve some form of intermediate steps where the corresponding phonological units contribute to achieving semantic access (ibid). In short, connectionist models "do not posit the eventual bypass of phonology", as phonological activation is intrinsic to achieving semantic access at all levels of reading development (Bjaalid, Høien, & Lundberg, 1997, p. 74). As Seidenberg (2007) noted, the difference between these models and earlier versions of the dual route lies in "the relative division of labor between the two pathways; [as] there is some input from both pathways for most words" (Seidenberg, 2007, p. 17).

Unlike the first and second explanations for the role of decoding in reading comprehension, the third explanation is more concerned with the *speed* with which the phonological code is obtained. According to limited capacity theories, humans can only execute a limited number of mental operations simultaneously when performing complex activities (Shiffrin & Schneider, 1977). Therefore, in the context of reading, a largely complex activity, some of the reading sub-skills would need to be automatized to free up resources for other demanding operations, such as comprehension (Perfetti, 1985). The choice of which operations to automatize is dependent on the nature of the operation. Decoding can be considered a good nominee for automaticity, as the mapping between symbol and sound is largely systematic (Koda, 2013). Empirical support for this assumption also exists. For example, research has shown that decoding speed was able to discriminate among L1 children's reading comprehension, even after controlling for decoding accuracy (Johnston & Kirby, 2006; Joshi & Aaron, 2000). However, evidence also exists for absence of additional contribution of decoding speed (e.g., Adlof, Catts, & Little, 2006), though it is possible that this was due to the use of an untimed reading comprehension measure (García & Cain, 2014).

Finally, a fourth explanation for the role of decoding in reading comprehension may also be at play. Particularly, it is possible that decoding skill contributes to real-time reading comprehension by modulating situational affect. Though this assumption has not been tested empirically, relevant findings in the literature seem to support it. On the one hand, it is interesting to note that links between affective constructs (for example, motivation, anxiety) on the one hand and decoding skill, on the other, have been demonstrated in the literature. Particularly, this relates to studies proposing the presence of low motivation among readers with poor decoding skill (Erler & Macaro, 2011); and also to studies proposing the presence of L2 reading anxiety in situations where the L2 and L1 writing systems are different (Saito, Horwitz, & Garza, 1999). On the other hand, links between affective constructs (for example, motivation, anxiety) and reading comprehension have also been demonstrated. Particularly,

this relates to studies demonstrating correlations between reading comprehension and affect (for example, "anxiety" in Zhao (2013), and "motivation" in Guthrie et al. (2007)). With the lack of causal evidence in this literature acknowledged (though see Sorg and Whitney (1992) for causal evidence on the impact of anxiety on lowering working memory functioning and task performance), it is not unreasonable to expect readers to experience negative feelings when encountering difficult-to-decode words, and to be disrupted by these feelings when reading for comprehension.

2.1.2.2.6.3.Empirical support

L1 decoding skill has been found to be a key predictor of L1 reading comprehension not only among children but also adults, and this contribution has been found to be independent of that of key predictors of L1 reading comprehension, such as vocabulary knowledge (e.g., Cunningham, Stanovich, & Wilson, 1990). As far as the L2 context is concerned, however, our knowledge of the contribution of L2 decoding proficiency to L2 reading comprehension ability is lacking (Nassaji, 2014; Pasquarella, Gottardo, & Grant, 2012), especially among adult L2 readers. The lesser attention to the role of L2 decoding is not surprising as, in the first place, only a small number of componential studies exist in the L2 context (i.e., studies that tested the contributions of multiple predictors of L2 reading comprehension ability) (Jeon, 2011; Shiotsu, 2010). However, even among the componential studies available, decoding has been largely ignored with the L2 literature being more concerned with the contributions of linguistic knowledge (traditionally viewed as breadth of vocabulary knowledge and grammar knowledge), and L1 reading comprehension (see Alderson's (1984) question and Bernhardt's (2000, 2005) model, section 2.1.2.1.2).

In fact, Jeon and Yamashita's (2014) meta-analysis clearly showed the scarcity of studies that examined the relationship between L2 decoding proficiency and L2 reading comprehension performance. The authors searched major databases applying a large time-frame (1979 to 2011), yet, their search only yielded 18 studies. Compare that with García and Cain's (2014) meta-analysis on the relationship between *L1* decoding proficiency and *L1 reading comprehension* which adopted similar search criteria, and a decade shorter time-frame (1989-2012), yet yielded 110 studies.

What is more, of the 18 studies in Jeon and Yamashita's (2014) meta-analysis, no study tested adult ESL readers residing in non-English speaking countries[1], as the majority of the

[1] One study, Shiotsu (2010) did recruit adult ESL readers in a non-English speaking setting (Japan). However, the study did not employ a decoding measure that precisely reflect Jeon and Yamashita's, and the present study's, definition of decoding, (i.e., the ability to obtain the phonological codes from graphical codes accurately and quickly) as it used matching tasks.

studies were on language-minority-children in English speaking countries. Needless to say, it is difficult to rely on findings from language-minority children, as this population is situated in the second language context and thus may be more exposed to auditory vocabulary than adult ESL readers, and particularly those in foreign language contexts.

Returning to Jeon and Yamashita's (2014) meta-analysis, one of the important findings that emerged from their study is that the L1WS modulated the size of the L2DP-L2RC correlation, with studies recruiting logographic ESL readers reporting a lower correlation (r= .36) than that found in studies recruiting alphabetic ESL readers (r=.57). This could be because ESL readers with logographic L1 backgrounds rely less on phonology in L2 reading than those with alphabetic L1 backgrounds (Hamada & Koda, 2010). Regardless of what the reason might be, this finding underscores the importance of examining the influence of L2 decoding proficiency to L2 reading comprehension across a range of L1WSs.

A handful of studies have moved beyond bivariate correlational evidence, adopting methods such as regression. To the researcher's knowledge, the earliest was Koda's (1998) study on Korean and Chinese beginner-level ESL readers. In this study, around 50% and 27% of the L2 reading comprehension variance was accounted for by the Korean and Chinese participants' L2 decoding proficiency, respectively. (L2 decoding proficiency was operationalised as the accuracy of pseudoword decoding using Woodcock's (1987) Word Attack Subtest). Examining readers from two L1WS backgrounds is a strength of Koda's study, as it emphasized the role of the L1WS in modulating the contribution of L2 decoding proficiency to L2 reading comprehension. However, the very small sample sizes (20 per group), along with other limitations (lack of control variables and decoding speed indicators) could jeopardize the validity of her findings.

Jeon (2011) found that L2 decoding proficiency (also operationalised as the accuracy of pseudoword decoding) explained 12% of the variance in intermediate-level Koreans' L2 reading comprehension. However, Jeon's study, like Koda's (1998), has concentrated on decoding *accuracy*. Given the theoretical importance of decoding speed on reading comprehension, a speeded decoding test is necessary to fully understand the contribution of decoding proficiency to reading comprehension (Florit & Cain, 2011). In addition, though Jeon (2011) included other variables, for example, vocabulary knowledge, the fact that L2 decoding proficiency was placed first in their hierarchical regression model made it difficult to ascertain whether L2 decoding proficiency would contribute to L2 reading comprehension ability over

and above the contributions of other known L2 reading comprehension contributors (Jeon, 2011 only reported the change in R^2 that the second entered variable (VK) was responsible for and did not mention whether L2 decoding proficiency emerged as a strong predictor in their subsequent models/steps). Additionally, given that L2 decoding proficiency is differentially related to L2 reading comprehension in readers with different L1WSs, findings cannot be confidently generalized to the wider L2 reader population, particularly to readers with non-alphabetic WS backgrounds.

Jiang, Sawaki, and Sabatini (2012) noted that the relationship between L2 decoding proficiency and L2 reading comprehension has rarely been investigated among readers with logographic L1WSs, and sought to examine whether L2 decoding proficiency would contribute to L2 reading comprehension for these individuals. Their regression analyses showed that L2 decoding proficiency (operationalized as the accuracy and speed of real word and pseudoword decoding using TOWRE) (Torgesen, Rashotte, & Wagner, 1999) accounted for 11% of the reading comprehension variance in Chinese intermediate-level learners of English. However, the fact that the contribution of L2 decoding proficiency was studied independently of other L2 reading comprehension contributors again makes it difficult to ascertain whether L2 decoding proficiency would contribute to L2 reading comprehension over and above the contributions of known predictors of L2 reading comprehension, for example, vocabulary and grammar knowledge, as noted by Nassaji (2014). Given that vocabulary knowledge, grammar knowledge, and L1 reading comprehension have been shown to correlate highly with L2 reading comprehension ability, and moderately with L2 decoding proficiency, L2 decoding proficiency might have contributed to L2 reading comprehension because proficient decoders are also readers with wider vocabulary sizes, greater grammar knowledge, and/or more proficient L1 reading comprehension skills. Simply put, variations in L2 decoding proficiency, which accounted for variations in L2 reading comprehension, might have reflected variations in L2 vocabulary knowledge, L2 grammar knowledge, and/or L1 reading comprehension skill.

Jiang (2016) tested the contribution of L2 decoding proficiency to L2 reading comprehension with decoding proficiency being assessed using measures of oral reading speed, accuracy, efficiency (words correct per minute), and prosody. Jiang's (2016) participants comprised four groups of adult ESL learners: Chinese, Japanese, Arabic and Spanish. Again, findings suggested that the relationship between decoding proficiency and reading comprehension was modulated by the L1WS, as indicated by the percentage of variance in reading comprehension accounted for by decoding (increasing from 18% to 32%, 46%, and

65% for the Chinese, Japanese, Arabic and Spanish groups, respectively). Jiang's (2016) study was particularly interesting as, to the researcher's knowledge, it appears to be the only study that tested the contribution of decoding to reading comprehension among Arab ESL learners. Nonetheless, Jiang's (2016) findings should be viewed with caution, as again a major weakness of his study is the lack of control variables.

Finally, although this section is concerned with reporting evidence from componential L2 studies, it is relevant to mention here available evidence from other research paradigms. For example, intervention studies provide a stringent test for the contribution of L2 decoding proficiency to L2 reading comprehension performance. In a recent large-scale study involving 600 English learners of French, Woore, Graham, Porter, Courtney, and Savory (2018) tested the impact of three programs of instruction on L2 reading comprehension: decoding (phonics) instruction (plus the use of challenging L2 reading materials), strategy instruction (plus the use of the same challenging L2 reading materials), and a program where participants were exposed to the challenging L2 reading materials only without any phonics or strategy instruction. Participants receiving the third intervention constituted the control group, and both the phonics and strategy groups were exposed to the same challenging texts that participants in this group were reading. Reading comprehension, among other variables, was assessed at three time points: immediately before the intervention, immediately after intervention, and 6 months post-intervention. Findings showed that, in all three groups, learners made statistically significant progress in L2 reading comprehension following the intervention program (Woore, Graham, Porter, Courtney, & Savory, 2018). However, no evidence was found for the superiority of any of the interventions in improving L2 reading comprehension. In other words, neither decoding (phonics) instruction, nor strategies instruction developed learners' reading comprehension over and above the provision of challenging and engaging texts.

At the outset, this may suggest that decoding instruction did not lead to significant gains in reading comprehension (given the absence of significant differences between the control and phonics instruction group). However, it is worth noting here that the control group received some level of intervention in the form of working with challenging and engaging texts which, as the authors noted, may have in itself improved learners' reading comprehension. The rationale behind including this group was that the inclusion of a typical control group (i.e., a group that simply follows usual French curriculum) would make it difficult to draw causal claims for the effectiveness of the strategy instruction. (This is because strategy instruction would inherently involve engaging with challenging texts, thus, making it difficult to pinpoint

exactly what led to reading comprehension gains: engaging with challenging texts - which provided rich learning opportunities - or strategies instruction). That strength in design, however, may have led to a saturation effect. That is, it is possible that within a 3-month period (the intervention period), learners' reading comprehension could only be improved to a certain limit that further intervention would fail to lead to additional noticeable growth. In other words, as the two intervention groups were already undergoing the intervention involving working with challenging texts, possible gains may have reached that maximum threshold point, potentially leading to the absence of a combined effect (i.e., the case where the decoding plus "challenging texts" intervention would lead to better reading comprehension outcomes than the "challenging texts" intervention on its own). This speculation is further supported by the finding that (1) all three groups improved their comprehension significantly following the intervention program. (Though one might argue here that some progress would be expected over 3 months of usual instruction, as the authors noted, prior evidence for limited progress among this population may be argued to eliminate this possibility); (2) and most importantly, the decoding group made the largest gains in vocabulary knowledge, as measured post-intervention. Given strong prior evidence for the role of vocabulary in reading comprehension (e.g., see Lesaux, Kieffer, Faller, & Kelley, 2010), it might not be reasonable to conclude that L2 decoding skill does not impact L2 reading comprehension.

2.1.2.2.6.4.Gaps in previous research into the contribution of L2 decoding proficiency to L2 reading comprehension

The review above has highlighted three key limitations of the literature into the contribution of L2 decoding proficiency to L2 reading comprehension, namely: (1) lack of control variables; (2) lack of decoding speed indicators; and (3) the seemingly L1-restricted validity. To these, three more limitations are added and discussed below: (4) use of single decoding measures; (5) employment of readers from a single proficiency level; and (6) lack of qualitative investigations.

Use of single decoding measures

Most L2 studies tended to use a single decoding measure (pseudoword *or* real word) and these were often used interchangeably as an index of readers' L2 decoding proficiency. However, it has been argued that the two measures may tap into slightly different underlying competencies. Though both assess decoding skill, a pseudoword measure places *more* emphasis on assessing readers' ability to obtain the phonology of orally-acquired words which they have not previously encountered in print, and also new words for which readers have no

prior phonological representations. Thus, it tends to reflect readers' ability to *compute* the phonology of words (Carr & Levy, 1990). Furthermore, where the pseudoword measure is timed/speeded, then it would also assess readers' ability to perform that computation quickly. A real word measure, on the other hand, places more emphasis on assessing readers' ability to decode mostly known words for which they have existing phonological representations, and is thus likely to assess the quality of these existing representations, along with readers' ability to recall them quickly (where the measure is timed/speeded). Thus, if a reader's decoding proficiency is assessed using pseudowords only, we risk not assessing readers' potentially existing phonological representations, nor their ability to access these representations quickly. Similarly, if we assess readers' decoding proficiency using mostly real (familiar) words, we risk not assessing readers' ability to compute the phonology of unfamiliar written representations, nor their ability to perform this computation rapidly.

Thus, only when researchers include both decoding measures are they able to: (a) ensure a thorough assessment of decoding skill; and (b) test whether these two measures would contribute to L2 reading comprehension independently of one another. However, it is important to note here that (1) learners may use previous word knowledge to decode pseudowords, (for example, using their knowledge of the word <thunderstorm> to read the pseudoword <thunderlip>, and may also compute the phonology of known words rather than retrieve them from their memory (which is particularly possible if the real word items were unfamiliar), both issues often raised in almost every manual of standard decoding tests (e.g., Torgesen, Rashotte, & Wagner, 2012); (2) the two skills tend to more or less develop in tandem, thus it is unlikely, though not impossible, to find a learner with excellent real-word decoding skills but very poor pseudoword decoding skills, and vice versa. These two issues could lead to the presence of a large correlation between these two measures (for example, see the figures reported in Torgesen, Rashotte, and Wagner (2012)). This large correlation might in turn make it difficult to statistically assess the unique contributions of these two measures. Having said that, however, it is worth noting that in their construct validation study, Torgesen, Rashotte, and Wagner (2012) found that, despite a high correlation between the two variables (r=.8), *and despite the real word measure including unfamiliar items*, the two skills loaded onto two different factors, suggesting that the two skills, despite being significantly related, remain different.

Recruitment of readers at a single proficiency level

Studies into the contribution of L2 decoding proficiency to L2 reading comprehension are often small-scale, recruiting readers from a single proficiency level. This inhibits opportunities to understand the developmental trajectory of the contribution of L2 decoding proficiency to L2 reading comprehension (i.e., whether the magnitude of the contribution of decoding increases, decreases, or stays constant throughout the course of L2 development)

Logic suggests that the contribution of L2 decoding proficiency may vary across readers at different levels of L2 proficiency. It is possible that variations in the decoding proficiency among a group of low proficiency readers (beginners) would contribute less to explaining variance in reading comprehension scores because even if decoding was proficient (accurate/fast), readers' overall low proficiency may mean that they lack knowledge of word meanings, grammar knowledge, etc. Consequently, any freed cognitive capacity may not be very beneficial as language proficiency is low in any case. Similarly, any accurate decoding may not be very beneficial when word meanings are unknown. On the other hand, variations in the decoding proficiency among higher proficiency readers (say at intermediate levels) may mean that some readers can access word meanings successfully, and others cannot; And some might have more freed cognitive capacity than others, which, given their high proficiency, could be beneficial for their reading comprehension. Thus, L2 decoding proficiency might be more predictive of L2 reading comprehension among readers at higher levels of proficiency than among readers at lower levels of proficiency. Simply put, L2 proficiency may have to reach a certain minimum threshold after which the contribution of L2 decoding proficiency may become significant.

An opposite trend, however, may also be present (i.e., that the contribution of L2 decoding proficiency would decrease as linguistic proficiency increases). Given their more sophisticated linguistic knowledge, readers with higher levels of proficiency (i.e., vocabulary and grammar knowledge) might be less disrupted by poor decoding than learners with lower levels of linguistic proficiency. This is because high proficiency learners are more likely to be able to compensate for any decoding deficiencies (for example, by being more skilled at inferring word meaning from the sentential context, thus being less disrupted by poor decoding, and/or more able verify the accuracy of their decoding realisations) (see the assumption of "compensatory processing" in reading, Stanovich, 1980, 1986, 2000). Thus, it can also be hypothesized that once linguistic proficiency reaches a certain threshold, variations in decoding proficiency may not lead to significant variability in reading comprehension performance.

However, given that readers from different proficiency levels were never investigated in a single L2 decoding study, this remains a speculation only. However, it is worth noting that in an attempt to answer this question, Yamashita and Jeon (2014) included an age moderator in their meta-analysis, and found that age did not moderate the relationship between L2 decoding proficiency and L2 reading comprehension. It goes without saying, however, that in second language settings, one's age provides few clues about one's linguistic proficiency, which Jeon and Yamashita (2014, p.198) later acknowledged:

> In the case of L2 reading, L2 proficiency, rather than age, may be an influential variable that moderates the relationship between L2 decoding and L2 reading comprehension.

Understanding how decoding skill might constrain reading comprehension in learners at different proficiency levels would have valuable theoretical and pedagogical implications. Theoretically, the presence of a moderation effect would suggest that the contribution of decoding to L2 reading comprehension is not stable, and that this should be taken into consideration in componential models of L2 reading comprehension. Pedagogically, such a finding would inform our understanding of the learning needs of readers at different proficiency levels.

Lack of qualitative investigations

The final limitation relates to the lack of "complementary" qualitative investigations. In common with much of the wider L2 reading comprehension research, as Alderson (2000) noted, studies into the contribution of L2 decoding proficiency to L2 reading comprehension have been predominantly quantitative, with knowledge being largely constructed from abstract test scores, which ignore the participant's voice. That is, researchers interested in investigating the contribution of L2 decoding proficiency to L2 reading comprehension would test the participants on these variables, determine the percentage of variance that L2 decoding proficiency has explained in L2 reading comprehension, and offer explanations for the contribution which often reflect an existing theoretical stance.

This purely quantitative approach to understanding the contribution of L2 decoding proficiency is undoubtedly valuable; however, it does have some limitations:

(1) given the correlational nature of the design, causal inferences cannot be directly made;

(2) this approach does not provide any clues into the mechanisms that underlie the contribution of the predictor of interest to the outcome variable;

(3) the possibility of type I or II errors cannot be completely dismissed (though see below); and

(4) though less importantly, it leaves unanswered the question of what might account for any unexplained variance in the regression model it tests.

The first of these limitations may be difficult to address at present as to be able to establish causation, an intervention design is needed. However, in the absence of strong observational/correlational evidence for the contribution of L2 decoding proficiency to L2 reading comprehension, it may not be appropriate to adopt an intervention design, at least at this state of knowledge (see Caruana, Roman, Hernández-Sánchez, & Solli, 2015 for argument against the use of intervention/longitudinal designs in the absence of strong observational/correlational evidence)

Having said that, it is worth noting that some analytical techniques (such as multiple regression) constitute powerful tools, as they allow the researcher to identify the independent variables responsible for any variance observed in the dependent variable on the basis of statistical significance (Tabachnick & Fidell, 2014). Further, the inclusion of covariates eliminates the effect of confounding variables and reduces error variance, i.e., variance caused by extraneous variables (Field, 2015). In fact, it has been argued that multivariate statistical techniques could guard against Type I errors 95% of the time (Tabachnick & Fidell, 2014)

The remaining three limitations (2, 3, and 4 above) provide a rationale for investigating the contribution of L2 decoding proficiency to L2 reading comprehension from a qualitative perspective. For example, interviews comprising task-based discussions may provide useful information as they could:

(a) act as explanatory devices for relations (Cohen, Manion, & Morrison, 2011), thus allowing an understanding of the mechanisms underlying the contributory role of decoding to L2 reading comprehension;

(b) validate quantitative methods of investigation (i.e., investigating whether decoding difficulties would emerge as one of the issues that interferes with learners' reading comprehension); and

(c) follow up on quantitative findings by enabling an understanding of the factors that might account for any unexplained variance in the regression model tested.

To date, however, participants' voices remain sidelined in prediction studies of L2 reading comprehension. This is despite the fact that: (1) qualitative components are increasingly being merged into correlational studies (e.g., see Kremmel, Brunfaut, & Alderson, 2015), hence the

emergence of Embedded Correlational Research Designs, in which a qualitative component is embedded within correlational designs with the intent of explaining the mechanisms that relate the predictor of interest to the outcome variable (Creswell & Plano Clark, 2007); (2) and most importantly, evidence is accumulating of readers' ability to reflect on what interferes with their performance on reading tasks. For example, in Erler's (2003) thesis, participants were able to verbalise how their decoding facilitates meaning access. As one participant puts it (p.167):

> " *If the teacher sometimes says [a] word, and some other time you [encounter the word] and don't know what it means, if you try and pronounce it in French and it will sound like what the teacher says then you might recognise it*".

Evidence for the "fast decoding- freed capacity" mechanism was also found. In post-reading-task interviews, one of Alsubaie's (2014, p.166) participants clearly verbalised how their slow decoding, caused by English irregular GPC, placed a strain on the limited capacity system:

> *"The words that have [irregular spellings] I spend time pronouncing them and forget what I have already read, from this point I start isolating the ideas*".

To sum up, qualitative investigations could complement correlational studies not only by providing triangulation of correlational findings, but also by providing explanations for these findings. These accounts may then validate, and potentially provide further insights to, our understanding of the theoretical mechanisms that underlie the contributory role of L2 decoding proficiency in L2 reading comprehension.

2.1.2.2.6.5.Conclusion and remarks

To conclude this section, very little progress has thus far been made to examine the contribution of L2 decoding proficiency to L2 reading comprehension. Correlational evidence suggests that the two variables are strongly related. The handful of studies adopting methods such as regression suggests that decoding makes a significant contribution to L2 reading comprehension. This evidence, however, is weakened by several issues: (a) a tendency to examine the contribution of L2 decoding proficiency independently of that of other, known predictors of L2 reading comprehension; (b) lack of decoding speed indicators; (c) the seemingly L1-restricted population validity; (d) limited operationalization of the decoding construct (i.e., use of single decoding measures: real word or pseudoword); (e) recruitment of

readers from a single proficiency level; and (f) lack of complementary qualitative investigations.

Now that the contribution of L2 decoding proficiency, along with that of other predictors of L2 reading comprehension, has been discussed, the next section reviews the literature on second language decoding acquisition, with a particular focus on the challenges encountered by second language learners as they proceed along the path of L2 decoding acquisition.

2.2. Second language decoding acquisition

At the outset, it should be noted that unlike other L2 skills (for example, vocabulary knowledge, grammar knowledge), little attention has been given to L2 decoding (Koda, 2005). As such, and to date, much ambiguity surrounds L2 decoding acquisition. Nonetheless, in the past three decades, efforts have been made to improve our understanding of L2 decoding. Much of the research in this area has been concerned with the specific *nature* of L2 decoding, and particularly how it differs in readers with different L1 writing system backgrounds (Nassaji, 2014; Cook and Bassetti, 2005). These attempts have enhanced our understanding of the linguistic challenges that L2 decoders encounter along the path of L2 decoding acquisition.

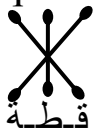
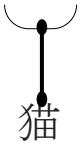
Given the centrality of the writing system in learning to decode (Perfetti, 2003), the review below begins by presenting the concept of a "writing system", along with a discussion of how writing systems shape the nature of the L1 decoding process. It then provides a relevant theoretical framework for the role of L1 decoding skill in L2 decoding acquisition. A description of the Arabic writing system as well as the decoding processes involved in its reading are also presented. Relevant empirical evidence from quasi-experimental, descriptive, and qualitative studies is then reviewed. The section concludes by highlighting gaps in previous research into the decoding challenges of L2 learners, leading to the aims of the present study and its research questions.

2.2.1. Writing systems: an overview

The term "writing system" (WS) has been used to refer to the design principles that underlie written language (Perfetti et al., 2007). The unit of language represented in visual signs is what determines the type of the system: signs in alphabetic WSs represent phonemes; in consonantal WSs, primarily consonantal phonemes; in syllabic WSs, syllables; and in logographic WSs, whole words (Coulmas, 1999).

Table 1. Illustration of the basic mapping principles in alphabetic English, consonantal Arabic, syllabic Japanese (Kana), and logographic Chinese for the word 'cat'.

Alphabetic	Consonantal	Syllabic	Logographic

/kæt/	/qitʌh/	/neko/	/maʊ̯/
			

Notes:
- The mapping between the Arabic phonological representation of 'cat' and the Arabic orthographic representation appears intersected because IPA is read from left to write whereas Arabic is read from right to left.
- As English is a deep alphabetic writing system (see below), a grapheme may comprise more than one letter, for example, the letter string <gh> in <cough> would map onto the single phoneme /f/.
- English, though primarily alphabetic, contains morphemic symbols, such as <ed> (Cook and Bassetti, 2005)
- Chinese has also been described as a morphemic writing system (Cook and Bassetti, 2005).

Apart from the linguistic unit represented, WSs also differ with respect to their orthographic depth, i.e., consistency in grapheme-phoneme correspondences (GPC). Some, described as shallow WSs, represent sounds following consistent GPC, while others, referred to as deep, exhibit unreliable GPC (Frost, Katz, & Bentin, 1987). For example, Italian and English are both considered alphabetic WSs, but they differ in depth, with Italian being shallow as it mostly exhibits a one-to-one GPC whereas English is deep, as the correspondences are less consistent, for example, <c> can be read as /s/ or /k/ (Cook & Bassetti, 2005). In addition to the linguistic unit of representation and orthographic depth, WSs also differ with respect to their physical properties, i.e., script and directionality (Cook & Bassetti, 2005). For example, Arabic is a WS that employs a right-to-left directionality; Russian is one that employs a Cyrillic script.

2.2.2. L1 writing systems and L1 decoding

L1 decoding studies have shown that the nature of the WS (mainly that of its representational unit and orthographic depth) can shape the decoding processes involved when reading that system. Thus, readers of different WSs will not exhibit the same decoding processes when reading their first language. These L1 processing variations have been the topic of two key theories: the Orthographic Depth Hypothesis (ODH) (Frost, Katz, & Bentin, 1987) and, later, the Psycholinguistic Grain Size theory (PGST) (Ziegler & Goswami, 2005).

The Orthographic Depth Hypothesis (Frost, Katz, & Bentin, 1987) uses the Dual-route framework (Coltheart, Curtis, Atkins, & Haller, 1993) to explain these cross writing system variations. Essentially, it argues that readers adapt their reliance on the two processing routes (the direct and indirect routes) depending on the depth of the writing system. The strong form

of this hypothesis claims that in shallow orthographies (such as Italian), phonological decoding occurs exclusively through the assembled phonology route (e.g., Bridgeman, 1987; Frost, Katz, & Bentin, 1987; Turvey, Feldman, & Lukatela, 1984). On the other hand, in deep orthographies (such as Chinese), readers are assumed to rely on the lexical route (though, as previously discussed, exploiting the lexical route should not imply "deactivation" of phonology, even in Chinese: see Perfetti's et al (1992) Universal Phonological Principle). However, empirical evidence – mainly derived from tasks testing the effects of word frequency – has repeatedly challenged this view. The rationale behind these tasks is that the effects of word frequency should be prevalent exclusively in the reading of deep orthographies due to their presumed reliance on the lexical route. However, these effects were also observed in the reading of shallow orthographies leading to the conclusion that both routes are activated in the reading of these orthographies. (e.g., Baluch, 1993; Baluch & Besner, 1991; Besner & Smith, 1992; Tabossi & Laghi, 1992, for Persian, Serbo-Croatian, and Italian, respectively). Thus, the revised, and the more tenable, form of this hypothesis states that across all WSs, both the direct and indirect routes are involved in obtaining the phonology of words, though to differing degrees (Katz & Frost, 1992).

The revised form of the Orthographic Depth Hypothesis provided a flexible framework for describing variations in decoding processes across WSs. Nonetheless it may be worth noting that Hebrew (and by extension, Arabic), though extensively referred to in the ODH, poses a small challenge to the hypothesis (Seidenberg, 1992). As Hebrew is typically written without short vowels, readers are not in actuality restricted to this two-route processing architecture because the sentential context is largely involved in the reading process. Seidenberg (1992, p. 102) concluded that "orthographic depth does appear to affect how Hebrew is read, though perhaps not in the manner suggested by the orthographic depth hypothesis. The effect seems to be realized in terms of the degree of reliance on contextual information in identifying [the phonological form of] words, rather than in terms of the degree of reliance on [single-word] phonology".

Moving away from the dual-route dichotomous structure, the PGST explains these cross-writing variations in a slightly different manner. The theory postulates that different 'psycholinguistic grain sizes' develop in response to variations in writing systems WSs ("grain size" refers to the amount of orthographic information necessary for phonological decoding (Cook & Bassetti, 2005, p. 16)). Put differently, this means that different grain sizes/decoding mechanisms will be employed across different WSs. To illustrate, according to the PGST,

readers of shallow WSs (for example, alphabetic German) can rely exclusively on small grain sizes, at the individual grapheme level, without making many decoding errors. On the other hand, in deep WSs, the unavailability (e.g., logographic Chinese), or inconsistency, (e.g., alphabetic English) of small grain sizes will result in the development of decoding mechanisms at more than a single grain size. That is, readers will also be required to learn correspondences at larger grain sizes, such as rimes (for example, referencing 'walk' as a basis for reading 'chalk') (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995) or even whole words, for example, 'choir' (Goswami, Ziegler, Dalton, & Schnieder, 2003).

Furthermore, the PGST also argues that understanding reading development is crucial for understanding subsequent skilled reading. That is, it does not treat early and skilled reading as two independent phases. In so doing, it gives proper weight to the role early reading development plays in structuring subsequent adult reading and its accompanying mechanisms. One example of a 'developmental footprint' that Ziegler and Goswami (2005) reported is the finding that skilled English and German readers show different length effects when reading words that are identical in both languages, for example, 'zoo-Zoo'. The German group were more sensitive to length indicating a decoding process that is more reliant on small grain sizes (Ziegler, Perry, Jacobs, & Braun, 2001). The authors concluded that this behavior has been developmentally established, i.e., in early reading. In addition to that, the PGST argues that learners of shallow WSs will acquire decoding skill more rapidly than readers of deep writing systems, a claim which has been supported by previous research (for example, at the end of Grade 1, Greek children read approximately 90% of real words correctly (Porpodas, Pantelis, & Hantziou, 1990). On the other hand, children reading English scored no more than 70% correct at the end of Grade 2, i.e., even a year later (Frith, Wimmer, & Landerl, 1998)).

With its introduction of the more flexible grain size aspect, the PGST provides an improved alternative to the dichotomous understanding of the Orthographic Depth hypothesis and an elegant perspective for describing the reading process across WSs. Nevertheless, 'universality checks' do remain a major challenge to any reading theory. In this regard, the PGST is not without its weaknesses. The theory has been critiqued for its lesser consideration of: fluency (e.g., Wimmer, 2006 in German), morphology (e.g., Durgunoğlu, 2006 in Turkish) the involvement of larger gain-size reading mechanisms in shallow WSs (Spinelli et al., 2005 in Italian), and sentential context (Alghamdi, 2015 in Arabic).

Now that the concept of "writing system" and its role in shaping L1 decoding have been presented, the next section presents L2 decoding acquisition along with a relevant theoretical framework.

2.2.3. L2 decoding acquisition: theoretical framework

According to the universal grammar of reading, reading has universal properties that are found regardless of the nature of the writing system, with the key one being that print/writing systems encode speech (Perfetti, 2003). This reading universal has implications to the second language (and particularly adult) learner. Since this understanding is not language-specific, once it is acquired in one (the first) language, it will be readily available and fully functional when they begin to decode the second language (Koda, 2008). Thus, first language decoding experience should provide direct facilitation to all L2 decoders, at least with regard to this general mapping principle (ibid).

However, being able to decode in any language necessitates an understanding beyond that basic reading principle. L2 decoders need to be fully familiar with *how* their L2 writing system encodes language, as well as the exact mapping details (ibid). To illustrate, for one to decode English, they need not only understand that print represents language, but also what elements of language are represented in print (for example, that letters represent mostly phonemes) and also the detailed grapheme-phoneme system (for example, that <n> represents the phoneme /n/) (ibid).

As discussed earlier, writing systems vary with respect to *ho*w they encode language in their visual signs, though many share similarities. It follows then that L2 learners do not stand on the same ground when beginning to acquire L2 decoding skill. That is, one group of learners may encounter more (and/or different nature of) difficulties along the road to L2 decoding acquisition. Thus, most L2 decoding research has concentrated on how L1-L2 WS distance can influence L2 word reading (Cook & Bassetti, 2005). In the following section, relevant theoretical background is presented along with relevant empirical evidence.

Before I proceed to this section, however, it is necessary to mention that factors additional to that of WS may be critical for L2 decoding acquisition (for example, environmental factors such as instructional methods and exposure to L2 print). Nonetheless, as the present thesis is more concerned with linguistic challenges, these environmental factors are not discussed here.

2.2.3.1.Orthographic Distance Hypothesis (Koda, 2008; Koda, 2013)

Koda argued that L2 decoding skill emerges from continual interactions between L2 print input and L1 decoding processes, and that L1 decoding processes will gradually transform themselves to be serviceable in the L2WS. WS distance – the degree of similarity/differences

between two WSs – largely determines the amount of transformation needed. When the two WSs induce similar processing demands, transferred L1WS decoding skill will be readily functional with only minimal (if any) adjustment. When the two WSs are different, L1WS processes must be substantially modified or they could interfere with the acquisition of L2 decoding skill. WS distance could therefore account for the disparate rates at which L2 decoding develop among L2WS readers with similar/dissimilar L1WS backgrounds, as the latter group is more susceptible to encountering challenges when decoding the second language.

Koda's model provided a framework that accounted for several findings in the L2 word reading literature (see section 2.2.5). However, it may be worth noting that, in her theoretical discussions, Koda (2013) often refers to two aspects of writing system distance: unit of representation and orthographic depth, and gives no mention to other writing system aspects that may potentially interfere with L2 decoding development. These could include (a) differential physical properties, such as script; (b) differential phonological inventories of the languages represented by the writing systems (for example, phonemic systems, syllable structure); and (c) differential L2 GPC mappings (for example, the grapheme <c> maps onto phoneme /k/ or /s/ in English, but onto /dʒ/ in Turkish). It seems reasonable to assume that similarity in physical properties, phonological inventories, and GPC mappings, could accelerate L2 decoding development. Furthermore, Koda's hypothesis is framed in a manner that assumes the "mono-literacy" of L2 learners, i.e., that L2 readers have developed reading processes in response to one WS only. However, in many cases, learners may acquire literacy in more than one WS, as is the case of the Arabic reader of English who is biliterate in two WSs, Alphabetic and consonantal (see section 2.2.4).

As the present study involved learners with Arabic L1WSs backgrounds, it might be helpful to describe the Arabic WS next, along with its decoding processes.

2.2.4. The Arabic writing system

Consonantal Arabic is, like English, a sound-based WS in that it connects graphemes with phonemes. However, the graphemes represented in print are primarily consonantal ones along with 3 long vowels which can also function as consonants (for example, the grapheme <و> can be decoded as /u:/ as in <توت> /tu:t/ <berries> , or /w/ as in <ولد> /wʌləd/ <boy>). Short vowels are not written as letters, but as diacritics <ً>, <ِ>, and <ُ>, for /ə/, /i/, and /u/, respectively. These diacritics (short vowels) are only implemented restrictively. Their use is limited to special texts such as the Quran, classical poetry, and children's books (Bauer, 1996).

Thus, normal texts are typically presented without short vowels. To illustrate, the word <pen>, pronounced as /qʌləm/, would typically be represented using consonants only <قلم> <qlm> (the short vowelled representation being /قَلَم/, with two diacritics added to represent the short vowels /ʌ/ and /ə/).

However, all children read a vowelled Arabic (which is a shallow alphabetic WS) in their first three years of schooling until they reach the fourth grade, where non-short-vowelled materials (i.e., the consonantal WS) are introduced (Abu-Rabia & Taha, 2006). As short-vowelled Arabic (i.e., the alphabetic WS) is a shallow WS with almost one-to-one GPC, Arabic readers could employ a small grain size decoding mechanism when reading words without making reading errors. However, to decode non-short vowelled texts (i.e., the consonantal WS), phonological information is generated through accumulation of information from several resources. As there is a one-to-one (consonantal) grapheme-phoneme correspondence, a small grain size decoding mechanism would suffice to decode consonants accurately. However, due to the absence of short vowels, readers also engage in an inference process which is aided by their morphological knowledge,[2] prior lexical knowledge, and the sentential context to derive the phonological information related to vowels (Taouka & Coltheart, 2004).

Furthermore, in terms of the phonemes represented by this system, modern standard Arabic (MSA) includes all the consonants found in English expect /p/, /v/, /g/, /ŋ/, and /tʃ/. The vowel system of MSA is, however, much smaller than that of English (see the "standard" English vowel chart below). For example, modern standard Arabic has six vowel sounds: three short ones /i/, /u/, and /æ/, and three long ones /i:/, /a:/, and /u:/ (Saiegh-Haddad & Henkin-Roitfarb, 2014; Saigh & Schmitt, 2012)[3]. It is also worth mentioning that modern standard Arabic is different from spoken Arabic (Ayari, 1996) and that varieties of spoken Arabic may have consonant sounds additional to those represented by MSA: for example, /g/ is found in the Arabic spoken in Saudi Arabia (Abdoh, 2011). Furthermore, most varieties of spoken Arabic, including Saudi Arabic, have two additional vowel phonemes: /o:/ and /e:/ (Abdoh, 2011;

[2] Arabic morphology is non-concatenative, i.e., morphemes are not ordered linearly, as in English. Words are composed of two bound morphemes; a consonantal morpheme, and a vocalic morpheme. The root is composed of consonants. This root cannot stand alone and is built on the other morphemic unit, the word (vowel) template. The word template is primarily composed of the short vowels, which are not represented orthographically in the unvowelled version of the writing system. High morphological awareness can in part (along with the sentential context and prior lexical knowledge) aid the reader to derive the suitable vowel template for the word (i.e., the vowel letters).

[3] Arabic presents an example of diglossia as the written and spoken varieties of the language are different (Abu-Rabia, 2002; Saiegh-Haddad & Henkin-Roitfarb, 2014). Modern standard Arabic (also referred to as "literary Arabic" and "written Arabic") is the language used for formal education (i.e., the language taught in schools) and is the one mostly used for written communication. Modern standard Arabic is sometimes described a "second language" for Arabic speakers as spoken Arabic (the language used for oral and everyday communication) is considered their native language given that it is the language they acquire orally prior to literacy instruction (Abu-Rabia, 2002; Ayari, 1996).

Saiegh-Haddad & Henkin-Roitfarb, 2014). Furthermore, /ə/ (or /ʌ/) is often observed as a variant of /a/ (ibid).

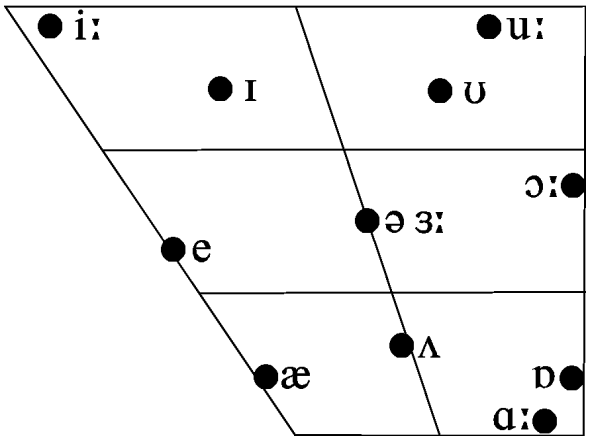


Figure 1. Standard English vowel chart

Finally, in terms of physical properties, both short-vowelled (alphabetic) Arabic and non-short vowelled (consonantal) Arabic employ a non-Roman script and a right-to-left directionality.

To sum up, Arabic and English are similar in that both are considered sound-based writing systems. Furthermore, short-vowelled Arabic and English are both alphabetic writing systems, though exhibiting different levels of orthographic depth, with the former being shallow and the latter deep, as well as different physical properties (i.e., script and directionality). Non-short vowelled Arabic (i.e., the consonantal writing system) is different from Alphabetic English in terms of its representational unit (primarily consonantal phonemes), orthographic depth (one-to-one consonantal GPC but more opaque vowel GPC given the absence of short vowel signs), and again its physical properties (script and directionality).

2.2.5. Empirical evidence

Koda's orthographic distance hypothesis (see section 2.2.3.1) has been formulated in response to studies reporting the decoding challenges encountered by L2WS readers. In their comprehensive review of L2WS reading research, Cook and Bassetti (2005) outline the research designs in this line of enquiry, and for organizational purposes, the studies reviewed below are arranged as per their respective designs. Furthermore, as English is the L2 of interest in the present work, most, though not all, of the studies included in this section recruited L2 English readers.

2.2.5.1. Quasi-experimental studies

2.2.5.1.1. L2 decoders with congruent versus incongruent L1WS backgrounds

In this body of research, proficiency-matched readers are placed in different groups according to a predefined criterion: their L1WS. A "congruent" group is one whose L1WS is similar to the target L2WS mostly in terms of the unit of representation (for example, Alphabetic L1WS and alphabetic L2WS). The "incongruent" group is one whose L1WS is different from that of the L2WS (for example, logographic L1WS and alphabetic L2WS). The groups are then asked to perform a reading-aloud task using word and/or pseudoword stimuli. Accuracy and/or speed of decoding data is then subjected to between-group tests with a view to see if prior L1WS processing experience impedes/facilitates L2 word reading.

2.2.5.1.1.1. Readers with Logographic L1WS vs. readers with alphabetic L1WS when reading an alphabetic L2WS

The majority of studies adopting quasi experimental designs recruited ESL readers from logographic L1WS backgrounds (for example, Chinese) and ESL readers from alphabetic L1WS backgrounds (for example, Korean) (Cook & Bassetti, 2005; 2008). It has been argued that logographic readers decode their L1WS by relying on a holistic decoding process, as 'words' cannot be broken down into letters (ibid) (Using a rough analogy, this is how you decode "**95%** passed math **&** science" as "**Ninety-five percent** passed math **and** science"). Because ESL readers might be influenced by their L1 processing experience, it was hypothesized that readers with logographic L1WS backgrounds might encounter more challenges decoding the L2WS in comparison with readers with alphabetic L1WS backgrounds (i.e., their decoding is likely to be more prone to errors and/or slower). This deficiency has been argued to be the outcome of their engagement in whole-word decoding that is insensitive to a word's component letters (Hamada & Koda, 2008). Such a process can be problematic, as readers would encounter challenges decoding new words/pseudowords whose phonology needs to be computed rather than obtained from memory, and also when decoding similar-looking real words which can be confused with one another (Grabe, 2009; Haynes & Carr, 1990).

Using reading-aloud tasks, Wang and Koda (2007) and Hamada and Koda (2008, 2010) compared the decoding proficiency (accuracy and speed of decoding) of advanced-level Korean and Chinese ESL readers based in the USA. English and Korean were deemed congruent as both are alphabetic WSs (i.e., similar unit of representation). Despite being matched on English reading proficiency, the Korean group outperformed the Chinese group in

speed and accuracy of decoding. The authors concluded that the differences observed are attributable to the participants' L1 processing experience. The processing demands imposed by their L2WS are similar to the demands of the Korean L1WS, meaning that L1-based processes had to undergo only minimal adjustments.

However, it is important to view the findings of these studies with caution. First and foremost, despite the researchers' efforts to ensure that the two groups are similar in terms of their proficiency level, several factors, other than the L1WS, may have contributed to these differences (for example, prior L2 instruction methods, amount of exposure to L2 print, age of the participants when L2 literacy started, knowledge of a third WS, etc.). Full consideration of possible confounding variables is necessary in quasi-experimental designs if one is to draw conclusions pertaining to the L1WS being facilitative or debilitative to L2 decoding acquisition. Obviously, however, one must acknowledge the presence of practical constraints here. That is, it is virtually impossible to ascertain that the sole distinguishing factor between two groups of learners is their L1WS. However, it might still be possible to improve the rigor of quasi experimental designs by collecting information about more of the possible confounding variables and controlling for these statistically.

Secondly, one weakness of these studies, which may be difficult to address, is that phonological, rather than WS, decoding challenges might have contributed to these differences. For example, a Korean reader may find it difficult to decode <freeze> because the Korean phonological inventory lacks the phoneme /z/. This is however not surprising as variations in phonemic inventories are often neglected in discussions of L2 decoding skill (e.g., see Koda's orthographic distance hypothesis, section 2.2.3.1).

Finally, it is worth mentioning that, in these studies, no analysis of readers' decoding realisations (outputs) is normally carried out despite the potential value of such analyses. For example, error-analyses of learners decoding outputs could: (a) provide explanations for the observed between-group differences (For example, it could help determine whether similar-word decoding errors were more prevalent in the logographic ESL group); (b) provide a thorough understanding of the nature of decoding difficulties experienced by learners with a given WS background.

2.2.5.1.1.2. Readers with consonantal L1WS vs. readers with alphabetic L1WS when reading an alphabetic L2WS

To the researcher's knowledge, no studies have yet been carried out to compare the ESL decoding proficiency of learners with consonantal vs. alphabetic L1WS backgrounds.

However, a few studies involving silent decoding tasks (i.e., word reading tasks that do not involve phonological production) are relevant to this discussion and thus merit reporting in this section.

Unlike logographic ESL readers' lack of attention to words' constituent letters, it has been hypothesized that consonantally-literate readers' difficulties would pertain particularly to *vowel* letters. Preliminary work was undertaken by Ryan and Meara (1991) when 10 Arabic ESL readers and 10 non-Arabic ESL readers – based in the UK and at intermediate levels of proficiency – took part in a same-different judgment task. The participants were presented with two presentations for each stimulus item. In the first presentation, they saw one correctly spelled word (for example, <department>). In the second, the word reappeared either in its correct spelling, or in a manipulated form in which a vowel was missing, (for example, <dpartment>). The participants had to decide whether the two presentations were identical. The researchers found that the Arabic group was slower and less accurate than the non-Arabic group, concluding that Arabic ESL readers "might be using mental representations of English words which ignore vowels and rely heavily on consonantal segments" (p. 539) and coining the oft-used term 'vowel blindness'. However, a major weakness of the study, and one which the authors acknowledged, is that it only included a missing vowel condition (no missing consonant condition), rendering it difficult to determine whether the low performance of the Arabic group in the vowel condition was due to difficulties in processing vowels, or letters in general.

Fifteen years later, this limitation was revisited in a study by Hayes-Harb (2006). In one task aimed at replicating Ryan and Meara's (1991), she asked 10 US-based intermediate-level Arabic ESL readers and 10 non-Arabic ESL readers to undertake a similar judgment task, this time including a third condition in which consonants were deleted. Contrary to Hayes-Harb's hypothesis, no interaction was found between the stimulus type (vowel/consonant) and the language group. Results were thus contradictory to Ryan and Meara's (1991) study. The author attributed the results to her small number of participants (10 per group) and the task implemented. Therefore, Hayes-Harb (2006) conducted another experiment with a larger –yet still small– sample size (15 per group) and a different task (letter detection). Participants were presented with texts, with their task being to detect and circle all the instances of one consonant (t) and one vowel (o). Overall, the non-Arabic group detected more vowels than consonants. For Arabic ESL readers, on the other hand, detection accuracy of consonants and vowels was similar, a finding which was seen to support the 'vowel blindness' claim.

Building on the two studies above, and given the fact that it is only the short type of vowels that are absent in the Arabic consonantal WS, Saigh and Schmitt (2012) conducted a study to see whether Arab ESL readers are less sensitive to English short vowels than long vowels. (Saigh and Schmitt's study did not include control groups but is nonetheless relevant to the discussion here). 80 sentences in which familiar target words (which contained a selected set of short vowels and long vowels) were presented to Arab ESL readers. Each target vowel was presented in three conditions: correct (for example, "We had dinner") incorrect (for example, "I need to imprave my English"), or missing (for example, "Sam was absnt"). The participants were able to notice errors with long vowels better than those with short vowels which, as the authors argued, demonstrates that reading short vowels can be significantly more problematic for readers with consonantal L1WS backgrounds. Their performance in the missing vowel condition was also significantly better than the incorrect condition which, as the authors suggest, challenges a literal interpretation of the widely held 'vowel blindness' idea proposed by Ryan and Meara (1991).

However, it is worth noting that in both Ryan and Meara and Hayes-Harb's studies, no baseline language measures were administered, making it virtually impossible to determine whether the observed differences/difficulties are explained particularly by consonantal processing experience or by variations in language proficiency level. The small sample size (10 to 15 per group) may also limit generalizability of the findings. Furthermore, reservations emerge regarding the participants comprising their control groups. Ryan and Meara provided no information about their language backgrounds. Hayes-Harb did provide information - (Portuguese (2), Vietnamese (2), Spanish (3), Japanese (3), Turkish (2), and Korean (3)) - but with such diverse language backgrounds, different variables may come into play. One variable that has been posited to be potentially influential is script (Brown & Haynes, 1985; Cook & Bassetti, 2005; Wang, Koda, & Perfetti, 2003). The majority of the participants in the control group were readers of L1 Romanized scripts. It is therefore unsurprising that the Arabic group exhibited the least accurate performance overall; indeed even their *consonant* detection accuracy lagged behind the non-Arabic group. As for the Saigh and Schmitt study, few inferences can be made regarding the debilitative impact of consonantal processing experience when no control groups were involved, as the observed difficulty might simply be developmental, (i.e., learners from other L1 backgrounds may also find English short vowels more problematic than long ones).

However, and aside from these methodological limitations, there are serious issues with the conceptual validity of the "vowel blindness" claim. Alghamdi (2015) argued that the hypothesis may have been, in the first place, built on inaccurate premises which may explain the somewhat irreproducible findings. To illustrate, by the time Ryan and Meara's (1991) seminal work on Arab ESL readers' "vowel blindness" was published, research on the role of short vowels in Arabic reading was almost non-existent[4]. Years later, several studies on Arabic reading began to emerge (e.g., Abu-Rabia, 1996, 1997, 1998, 2001). This line of research has consistently shown that the presence of short vowels had three main effects on L1 Arabic reading: (a) it improved reading accuracy; (b) it improved reading comprehension; (c) it slowed reading speed. It is not clear why the presence of short vowels slows speed, but it is possible that this is because readers have to process diacritics posted above and under the letters which has been argued to be a cognitively demanding task that necessitates more attention and eye fixations (Abu-rabia, 2001). Nonetheless, this is not important here. What is important is that improved reading accuracy and comprehension, and a slower speed, are all indications that, when present, short vowels were not ignored, as their presence significantly affected reading. Furthermore, in another study, Abu-Rabia (2001) asked his Arabic participants to read Arabic texts which included incorrect short vowels (diacritics). The results showed that readers' accuracy and comprehension were adversely affected by the incorrect vowels (Abu-rabia, 2001), again suggesting that when present, short vowels were not ignored, as their presence influenced reading accuracy and comprehension. Despite this literature, however, the remaining two "vowel blindness" studies, Hayes-Harb (2006) and Saigh and Schmitt (2012), seem to have taken the results of Ryan and Meara's work as their point of departure and overlooked this L1 Arabic reading literature. In other words, the body of research on Arabic L1 reading has not been made use of in a way that would inform hypothesis building in the L2 reading context. As these L1 Arabic studies produced these findings (i.e., that whenever present, short vowels were processed), it seems illogical to postulate that Arabic ESL readers would face difficulties perceiving English vowel graphemes.

Secondly, even if one was to ignore this Arabic reading literature, there is another reason why the vowel blindness hypothesis is conceptually problematic. As mentioned earlier, all L1 Arabic readers learn to read an alphabetic writing system in their first years of schooling (and they may even continue to read that system later on when reading special texts such as

[4] Though see Roman and Pavard's (1987) early work who observed Arabic readers' eye movements when reading short-vowelled and non-short-vowelled texts and found that readers demonstrated more fixations, and were slower, when reading the short vowelled materials which suggests that, whenever present, short vowels were not ignored.

classical books, poetry, the Quran, etc.). Thus, Arabic readers were taught to read, and *are biliterate*, in two writing systems: alphabetic and consonantal. As such, it is strange that previous research dismissed the potential impact of this alphabetic processing experience in discussions of "vowel blindness". From a theoretical standpoint, and as per the PGST (see section 2.2.2), this alphabetic processing experience is expected to leave a 'developmental footprint' and, therefore, its role should not be overlooked. In fact, empirical evidence has shown that logographic ESL learners who were exposed to a second writing system[5] (alphabetic Pinyin, a Romanized Chinese WS used to facilitate reading Chinese characters in childhood) were more able to read English pseudowords than Chinese readers who were not exposed to Pinyin during the initial stages of L1 literacy development (Holm & Dodd, 1996).

Finally, it is important to include findings from a recent study (which was published at the time this thesis was being written) which provided a strong test for the vowel blindness claim. Using eye-tracking, Alhazmi, Milton, and Johnston (2019), who also appeared to have overlooked the L1 Arabic reading literature, wanted to see if Arabic ESL readers (n=30) ignore vowels when reading English words compared to native English speakers. Findings suggested that the Arabic group were slow, taking double the time to process English words. However, and contrary to the authors' expectations, both the Arabic and native English speaker group paid equal attention to consonants and vowels. The authors concluded:

> The principal conclusion of this study is that eye-tracking results contradict the assumptions … which underlie the vowel blindness hypothesis. Arabic readers do not seem to avoid vowels in reading English... Both native English speakers and the Arabic speakers in this study, broadly speaking, appear to pay equal attention to vowels and to consonants while reading. But the Arabic speakers in this study do look at more letters while reading, and spend longer on each word…This confirms, however, an important assumption about such learners, that they suffer from word-level processing problems (p.244).

Thus, the findings of Alhazmi, Milton, and Johnston (2019) challenged the "vowel blindness" assumption, and at the same time raised questions about the nature of the word reading difficulties of this group of learners.

2.2.5.1.2. Quasi-experimental studies: conclusion and remarks

To sum up the findings in this section, quasi experimental studies suggest that L2 readers, particularly those who encounter a new L2WS, may encounter difficulties reading their second language due to variations in the L1 and L2 WSs, and particularly those relating to the unit of representation. Nonetheless, and at least as far as Arab ESL readers are concerned,

[5] A second writing system for a given language (as opposed to a 'second language writing system') is defined as any supplementary WS for representing the same L1 language (Cook & Bassetti, 2005, p. 26).

the impact of this variation should be viewed with caution given the considerable weaknesses in the reasoning underlying the vowel blindness claim, as well as the methodological limitations of the studies available.

Quasi-experimental studies are useful in that they showed how WS distance could impact L2 decoding development. However, as Bassetti and Cook (2005) noted, "it is not clear that much mileage can be gained from continuing to make L1-to-L2 transfer a main theme of research (p.45)". The researcher here reiterates Bassetti and Cook's remarks particularly because several basic, but essential, issues remain ambiguous about L2 decoding acquisition and development.

Firstly, *the extent to which* L2 learners encounter difficulties with L2 decoding is not entirely clear. This ambiguity, however, is unsurprising given the objectives of previous quasi-experimental studies. Researchers often conduct experiments in light of preset hypotheses regarding the decoding difficulties of a particular group of learners (with difficulties often predicted on the basis of contrastive L1-L2 writing system analysis), and their interest is mainly centered around validating these hypotheses (i.e., whether the decoding proficiency of one group of learners would be higher than that of another group). Thus, driven by these objectives, previous quasi-experimental studies tended to assess decoding ability using idiosyncratic measures of decoding (i.e., a sample of words/pseudowords often prepared by the researcher) as opposed to using standard and comprehensive measures of decoding skill (as the goal is more to test the superiority of one group over another than to see what the performance of either group translates to in terms of *overall decoding ability*). However, that one group outperforms another tells us little, if anything, about the decoding ability of L2 learners. For example, one group might outperform another but, in terms of overall decoding proficiency, might exhibit what amounts to 20% mastery of decoding skill. Similarly, one group might underperform another but exhibits what amounts to 95% mastery of decoding skill. Thus, it is only when established and comprehensive measures of decoding skill are used, and decoding mastery is quantified for L2 learners at different developmental stages, that researchers will be able to better understand whether, and the extent to which, L2 learners encounter difficulties decoding their second language. Such understanding is important not only for pedagogy, but also for theory. For example, it would allow the field to understand several issues pertaining to L2 decoding development, e.g., what level of L2 decoding skill L2 readers would typically exhibit throughout the course of L2 development, how far in the decoding mastery spectrum they can go, and the stage at which they acquire the GPC of the second language, which as

Erler and Macaro (2011) noted, remains unclear.

Secondly, prior quasi-experimental studies paid little attention to achieving a direct understanding of learners' decoding difficulties. That is, the nature of L2 decoding difficulties is often speculated on prior to conducting the study. However, one group might outperform another, but that superiority might be caused by reasons other than those (and/or additional to those) hypothesized by the researchers. Without direct investigation of learners' decoding difficulties, researchers might risk developing an inaccurate or, at best, incomplete understanding of the difficulties that L2 learners encounter when attempting to decode the second language.

2.2.5.2.Descriptive studies

Unlike quasi-experimental studies, descriptive studies tend to examine a single group of L2 readers. In such studies, researchers may describe the extent to which readers are able to perform on decoding measures so as to understand if they encounter difficulties decoding their second language. Decoding outputs/realisations (i.e., pronunciations of words) are then phonologically transcribed and examined for errors. The decoding errors found can then be categorized (for example, a consonant-related error), and described (i.e., what realisations were produced for the problematic consonant). Contrastive analysis of the L1 and L2 WSs could also inform the analysis and discussion process here (Cook & Bassetti, 2005).

For example, Woore (2016) noted that while French and English are both alphabetic WSs sharing the same unit of representation, the two WSs exhibit different orthographic depths (with English being deeper than French) which may pose challenges to the English learner of French. That is, if these readers employ their L1 decoding processes, then their decoding performance might suffer as a result of their engagement in a large-grain-size decoding process (despite the similarity between the two WSs being undoubtedly facilitative in some respects). In Woore's (2016) study, below-intermediate French L2 readers took part in familiar and unfamiliar reading-aloud tasks. Despite readers' familiarity with stimuli in the former group, decoding accuracy was at 41.6%. The figure went down to 33.2% in the unfamiliar word list (despite the fact that this list contained items comprising the exact same spelling bodies chosen for the familiar word list), indicating the difficulties that L2 readers encounter with decoding the L2WS. To better understand the nature of the participants' decoding difficulties, Woore (2016) then coded their decoding realisations in categories, for example, 'correct', 'whole-word errors', etc.

As far as learners with Arabic backgrounds are concerned, there is scant research into the word reading difficulties of this group (Emery, 2005). A few studies were, however, found

which did not focus on decoding difficulties per se but which, nonetheless, included some descriptive data of the decoding difficulties of these learners. For example, Alsulaimani (1990) argued that Arabic readers of English will tend to "regularize" English words because "Arabic orthography is highly phonemically regular" (though see section 2.2.4), and, as such, that their decoding performance will resemble that of surface dyslexics. He carried out experiments to analyse the reading errors of English native speaker dyslexics and UK-based Arabic ESL readers (the latter comprised two sub-groups: beginner ESL readers (School pupils aged 11.8-18) and advanced ESL readers (PhD students aged 24 - 43). Findings suggested that the nature of the decoding difficulties of the Arabic and dyslexic groups were similar. Both groups attempted to (a) produce "regularized" pronunciations of words (for example, reading <break> as /bri:k/); (b) omit, delete, or transpose letters; (c) exhibit difficulties with consonant clusters; and (d) confused some graphemes (for example, p and b). Further, both the beginner and advanced Arabic groups exhibited similar nature of errors, though the beginner group made more decoding errors.

Alsulaimani's findings are helpful in that they showed what decoding difficulties this group of learners encounters. However, reservations should be made regarding claims of L1 transfer, similarity in the nature of the errors observed in his two proficiency groups (beginner and advanced ESL readers), and regarding the linguistic heterogeneity of his Arabic participants. First, it is difficult (and of limited use) to assume that an error is indicative of L1 transfer when learners from other language backgrounds are not tested, as that error could simply result from lack of L2 GPC knowledge. In fact, young beginner readers of L1 English have been shown to produce reading errors which share similarity with that made by dyslexic readers (Matterson, 1983); thus the errors observed may simply be developmental (i.e., beginner decoders, regardless of their L1 backgrounds, will tend to produce regularized pronunciations of words). Second, it is difficult to make assumptions about similarity in the nature of errors observed across the beginner and advanced group as the huge age difference between the two groups may contaminate the findings obtained. Third, Alsulaimani's Arabic participants came from different Arabic countries; hence, there is a range of "first languages" and phonemic inventories (note that MSA is often described as a second language) which could impact decoding performance (for example, it is not uncommon for an Egyptian learner of English to decode <the> as /zə/, as Egyptian spoken Arabic lacks the phoneme /ð/, but errors of this nature almost never occur among Saudi (and Gulf) learners of English, as /ð/ is part of the phonemic

inventories of the spoken Arabic in that region). This linguistic heterogeneity could pose some challenge to the reliability of Alsulaimani's findings.

As part of her thesis work, Emrey (2005) categorised the reading errors of 31 upper intermediate ESL Emirati readers into 3 categories: vowel errors (65% of total errors observed), consonant errors (28% of total errors observed), and phoneme inversion errors (7% of total errors observed). As an extensive analysis of these errors was not the focus of Emrey's (2005) thesis, the researcher did not present in detail what problematic consonants/vowels were found under each category, nor a detailed examination of the phoneme inversion cases. However, the researcher did mention that her learners encountered difficulties with consonant clusters (for example, reading the pseudoword <hermantic> as /hermæn**ə**tik/), and several L2 GPC rules (for example, the most frequent realisation of <jid> was /jaid/ suggesting that the participants were unaware of the "magic e" rule, i.e., that <i> would mostly – though not always – be lengthened and pronounced as /ai/ if the word ended with <e> (ibid)). This last observation is interesting because the fact that /ai/, rather than any other random vowel, was the most frequent realisation for /i/ could suggest that learners' vowel reading difficulties are likely caused by their lack of L2 GPC knowledge rather than "vowel blindness".

Finally, some findings from Alsubaie's (2014) thesis are worth adding in this section. Alsubaie's study aimed to investigate the reading comprehension challenges faced by Saudi English major students. He asked 16 students to verbalise their thoughts while reading aloud a text for comprehension. Alsubaie commented that "Challenges with decoding, particularly of long words, [was] a major source of incomprehension and reading challenges" (p.151). Furthermore, decoding errors constituted the largest single category of error with 37% of total errors observed (other categories being semantics, syntax, punctuation, and other difficulties). He further categorized the decoding errors into four types – Vowels (44% of total errors observed), Consonants (38%), Syllables (16%) and Whole-words (2%). (Alsubaie included vowel omission and insertion in the vowel category). Alsubaie (2014) did not conduct a thorough examination of the decoding errors observed while learners where reading aloud, though he provided samples for errors found in each decoding category along with pertinent remarks. These were found helpful and particularly relevant to the present study and are thus presented in detail below:

- Vowel category: Alsubaie noted that "many incorrect decoding of vowels were the outcome of students sounding out the name of the letter, rather than its phonic sound" (p.156) (for example, decoding <a> <private> as /praiv**ei**t/, <e> in <men> as /mi:n/, <i> in <illegaly>

as /aili:gæli/, <o> in <freedom> as /fri:do:m/, and <u> in public as /pju:blik/). Alsubaie mentioned that these errors may have resulted from current ELT practice in Saudi where phonics is ignored and alphabets are taught by letter names. These decoding errors were interesting to the researcher because again they suggest that learners were generally able to "see" the vowel letters, but nonetheless were unable to accurately convert these vowel graphemes to their corresponding phonemes, possibly due to lack of L2 GPC knowledge.

- Consonant category: This category included consonant graphemes and silent letters. Alsubaie provided examples for problematic consonants and these were: (1) <g> (e.g., <**g**uilty> decoded as /ʤilti/), (2) <p> (e.g., <**p**ublic> decoded as /bublik/), (3) <q> (e.g., <**q**uestion> as /kju:stʃin) (though this error may reflect difficulties decoding the following vowel <u> rather than the consonant grapheme <q>), (4) <r> (e.g., decoding <arrested> as /ərrestid/ (though this error may reflect difficulties decoding double consonants rather than the grapheme <r> per se), and (5) <v> (e.g., decoding <lea**v**e> as /li:f/). Because the problematic consonants mostly represented sounds that did not exist in Arabic, the author commented that incorrect decoding of consonants was mostly the outcome of Arabic lacking these consonant sounds (i.e., variations in the phonemic inventory of English and Arabic).
- Syllable-related error: This category included errors where a word was divided into syllables and where the syllables were decoded individually rather than blended together when producing the final phonological form, potentially indicating poor phonological awareness (blending) (Ehri, 2015). Syllable-related errors also included errors where entire syllables were added to existing words (for example, the syllable <es> was added to the target word <speech> resulting in the realisation <speeches>. Alsubaie noted that syllable-related errors led to changes in tense (for example, from present to past) and quantity (for example, from singular to plural).
- Whole-word errors: this category reflected instances where entire words (or parts of words) were wrongly decoded (for example, decoding "constitution" as "constation").

Though Alsubaie's thesis was not particularly concerned with decoding difficulties, his examples of reading errors provided some interesting insights about his learners' decoding difficulties. For example, his data suggest that: (1) Saudi (Arabic L1) ESL readers do not only encounter difficulties decoding English vowels, but also consonants and multi syllable words (note that quasi experimental studies were mostly concerned with vowel-related word reading difficulties thus potentially creating a limited understanding of the decoding difficulties

encountered by these learners); (2) vowel-related errors were mostly the outcome of learners pronouncing vowel letter names, thus mostly suggesting lack of vowel grapheme-phoneme knowledge, rather than "vowel blindness"; (3) consonant-related errors mostly occurred where the consonant sound was not present in the learners' L1, thus suggesting that writing-system distance at the level of the phonemic inventories can impact L2 decoding development. However, it is important to highlight here that Alsubaie's participants were English major students and as such their English proficiency may lie toward the higher end of the proficiency scale. As such, his findings may not be representative of low proficiency Saudi ESL learners.

2.2.5.2.1. Descriptive studies: conclusion and remarks

To sum up the findings in this section, descriptive decoding studies showed that L2 readers encounter difficulties decoding their second language and that these difficulties may be explained by writing-system distance (particularly variations at the level of orthographic depth (e.g., Woore, 2016), phonemic inventories (e.g., Alsubaie, 2014), lack of L2 GPC knowledge (e.g., Emrey, 2005), and poor phonological blending (e.g., Alsubaie, 2014)).

As far as Arabic ESL readers are concerned, the few studies available seem to suggest that these learners encounter various decoding challenges that extend beyond those relating to vowel reading. These include consonants (particularly those representing sounds that do not occur in Arabic), consonant clusters, multi-syllable words, as well as a tendency to generate "regularized" pronunciations of words. All in all, these descriptive studies appear to provide a better understanding of decoding among Arab ESL readers', though limitations relating to the proficiency level of the learners involved (i.e., mainly higher end of the proficiency scale) raise questions about the applicability of the findings to other learner groups (for example, low proficiency learners).

The strength of descriptive studies is that decoding difficulties are allowed to emerge from the data, rather than deduced a priori as is the case in previous quasi-experimental work. In so doing, they allow for a relatively direct and potentially more comprehensive investigation of learners' decoding difficulties. However, a few issues merit highlighting for descriptive studies to achieve a better understanding of the decoding difficulties of second language learners:

First, prior studies were often small-scale including participants from one proficiency level. This does not allow us to see whether learners at different levels of proficiency would exhibit similar/different decoding difficulties, an investigation which could provide rich information about the course of decoding development in L2 learners.

Second, one limitation of "error-analysis" is that participants may resort to "avoidance" when encountering difficult tasks (Ellis, 2008). In the context of decoding, this means that participants may avoid decoding items which they find difficult. Consequently, some decoding difficulties may go unreported in descriptive studies. Thus, it might be necessary to complement error-analysis data with data from other methods (for example, qualitative data) so as to ensure comprehensive understanding of learners' decoding difficulties.

Finally, descriptive studies typically show us what decoding difficulties learners encounter (for example, difficulties decoding the grapheme <p>) along with a description of the realisations observed for these problematic graphemes. Though this data is undoubtedly valuable, applying additional analytical techniques is likely to yield further useful information. For example, instead of just describing the realisations observed, additional computing of the "relative frequency" of these realisations might reveal some interesting patterns. Unfortunately, such analysis of decoding realisations is somewhat laborious and is not typically seen in descriptive studies (though see Woore (2013)). However, the presence (or lack thereof) of patterns could contribute to, and enrich, discussion of these decoding difficulties. To illustrate further, imagine that <b> emerged as a problematic grapheme in a dataset of decoding outputs. Examining the relative frequency of the realisations produced for <b> could provide an indication as to why it was problematic. For example, if readers consistently decoded <b> as /d/, but not as /s/, then this pattern of error might suggest difficulties with mirrored letters, and hence, the script of the L2 writing system. In the case of Arab ESL readers, for example, the presence of patterns in learners' realisations of vowel graphemes could provide insights into the "vowel blindness" claim. By way of illustration, Ryan and Meara (1991) argue that Arabic speakers might ignore vowels and rely heavily on consonantal segments. If this is the case, one would expect to find readers' realisations of a given problematic vowel to be random and lacking in consistency. On the other hand, consistency in the nature of the vowel errors observed could challenge (at least) a strong version of the vowel blindness claim (i.e., that Arab learners are not attentive to English vowels).[6]

2.2.5.3.Qualitative studies

Very few studies used qualitative methods when studying L2 decoding (Cook & Bassetti, 2005). Of the studies available, methods such as interviews and think-aloud protocols (e.g., Erler, 2003; Woore, 2010) were used.

At a macro level, getting readers to think-aloud as they perform decoding tasks could show

[6] Nonetheless, it should be noted that they do not challenge a weak version of the vowel blindness claim (i.e., that learners are less attentive to English vowels compared to readers of non-Arabic backgrounds), given the absence of learners from these groups in the present study.

what decoding processes they engage in in general (for example, breaking words into parts), including any involvement of L1 decoding processes (though caution should be exercised when discussing the latter, given the known limitations of think-aloud protocols, especially in the absence of control groups (see section 3.4.2.3). At a micro level, these think-aloud verbalisations could show what exactly in the decoding process had contributed to the production of inaccurate phonological outputs (for example, applying a small-unit processing mechanism where a large-unit mechanism is necessary). Thus, such investigation could provide further insights into the nature of difficulties encountered by second language decoders.

For example, Erler (2003) asked English readers of L2 French to reflect on how they approached words in a decoding task. The main finding of her study was that that readers were consistently influenced by the GPC patterns of their L1WS when decoding French words. Similarly, Woore (2010) set out to investigate the decoding processes employed by 12 English learners of French as they tried to decode French words. His participants engaged in a range of processes which were coded as follows:

(a) appeal to declarative knowledge of L2 symbol sound mappings (for example, where a participant made a comment about how a given French grapheme should be decoded);

(b) appeal to declarative knowledge of L2 general patterns (for example, where a participant made a comment about more general patterns of French GPC rules);

(c) Making explicit comparisons with the L1 (English) decoding system (for example, where a participant makes a comment about how a grapheme is realized in English)

(d) using analogy (for example, where a participant used the phonological form of a known written word to assist them in decoding an unknown word);

(e) breaking the target item into syllables (i.e., where a participant focuses on individual syllables);

(f) breaking the target item into individual graphemes (i.e., where a participant focuses on single graphemes); and

(g) subvocalizing possible pronunciations (i.e., where a participant attempts to produce intermediate vocalisations for the target items).

Viewed from a dual-route processing perspective, Woore's (2010) findings suggest that learners were actively employing the indirect route to read the target words, as evident in their attempts to break words into components and their active engagement in various problem-solving tasks in an effort to generate the final phonological forms (for example, by means of referencing their L1 or L2 GPC knowledge base, analogy-making, etc.). What is interesting

here is that, due to lack of secure L2 decoding knowledge, the processes above were often executed unsuccessfully (for example, referencing an incorrect L2 GPC knowledge base, drawing incorrect analogies, inadequate breaking up of words into syllables, etc.) which eventually contributed to the production of inaccurate phonological forms.

To the researcher's knowledge, no qualitative study was carried out to investigate the decoding processes of Arabic, and particularly Saudi, ESL readers. This is despite the fact that such an investigation could show what processes these learners engage in (including whether they show a tendency to ignore vowels), while also elucidating what contributed to their inaccurate decoding outputs. However, as part of his work into the reading comprehension difficulties of Saudi English major students, Alsubaie (2014) asked his participants to think-aloud while reading aloud a text for comprehension, and then interviewed these students to further investigate the reading comprehension difficulties that emerged from the data. Based on his observations of their reading aloud performance, Assulbaie commented that:

> Students struggled enormously with the decoding of both vowels and consonants, but also of syllables and entire words. This often led to the substitution of a word by an entirely different meaning, or of a verbalization with no meaning in English (p.160).

Further, when the author asked students about their wrong decoding attempts, 12 out of 14 participants said that they struggled with decoding issues. He further commented that none appeared to be aware of several L2 GPC rules, as exemplified in the comment of one participant who mis-decoded <social> as /soʊ.kəl/: *"I thought it is 'sokal' …I don't know that 'ci' sound as 'sh"*. Further, he commented that some participants verbalised difficulties decoding consonants representing sounds that do not exist in Arabic, as shown by the comment of a participant who encountered difficulties decoding the grapheme <p> "*I have tried many times to correct myself but you know we don't have such sounds in Arabic*".

2.2.5.3.1. Qualitative studies: conclusion and remarks

To conclude, qualitative studies, employing methods such as think-aloud protocols and interviews, could provide further insights into the decoding challenges encountered by second language learners. However, a qualitative approach to studying word reading is not without problems (For a full discussion of these limitations, see section 3.4.2.2 and 3.4.2.3). Therefore, employing additional research techniques (e.g., analysing learners' decoding realisations) may be necessary to ensure a fuller understanding of the decoding difficulties experienced by second language learners.

2.2.5.4. Conclusion and gaps in previous research into the decoding challenges of second language readers

Taken together, the key finding of past quasi-experimental, descriptive, and qualitative decoding studies is that L2 readers encounter challenges decoding the second language. As far as language/WS factors are concerned, these difficulties appear to be the outcome of both lack of L2 decoding knowledge and L1-L2 writing system distance, particularly at the level of unit of representation (e.g., Hamada and Koda, 2008; 2010), orthographic depth (e.g., Woore, 2010), and phonemic inventories (e.g., Alsubaie, 2014), providing support for Koda's (2008) Orthographic Distance Hypothesis (though note that Koda's theoretical work mainly discusses the "representational unit" and "orthographic depth" aspects, and gives no mention to distance at the level of phonemic inventories, or physical properties).

However, given the impact of prior L1WS processing experience on L2 decoding acquisition, it is important, for theory and pedagogy, to diversify the language groups studied in this line of research. Particularly, decoding studies involving readers with Arabic L1 backgrounds are scarce in the literature. From a theoretical standpoint, studying a limited group of L1s may restrict a full understanding of the difficulties that L2 learners encounter when decoding the second language. As an example, Bassetti and Cook (2005) mentioned that it is unclear if variations in L1-L2 physical properties could impact L2 decoding acquisition (Indeed, as noted in the above paragraph, most of Koda's theoretical work (e.g., see Koda, 2013), places emphasis on two aspects of WS distance: the unit of representation and orthographic depth). Clearly, a better understanding of L2 decoding acquisition will only be possible if the decoding challenges of learners from a variety of L1 backgrounds are surveyed. Further, from a pedagogical standpoint, the impact of L1WS on L2 decoding suggest that learners from different L1WS backgrounds may experience different hurdles along the path of L2 decoding acquisition. In practice, this means that the L2 decoding literature, as it currently stands, may not be optimally useful to certain language groups. An inclusive body of literature would provide a useful resource for policy makers, curriculum developers and ESL practitioners in various learning contexts.

Furthermore, the L2 decoding field would benefit from several methodological modifications. These include:

(a) the use of comprehensive and widely recognized decoding measures when assessing decoding skill. This would not only ensure valid and reliable assessment of decoding skill, but

would also facilitate comparability of research findings, particularly if the field is to begin to quantify the decoding ability of L2 readers throughout the course of L2 development;

(b) more direct investigation of L2 decoding difficulties, as a priori assumptions of the decoding difficulties of learners may be incomplete and/or, worse, inaccurate;

(c) thorough analysis of learners' decoding realisations including an analysis of how learners realise problematic graphemes, and the relative frequency of these realisations;

(d) employing more than one research method to study the decoding challenges of L2 learners;

(e) inclusion of learners at different levels of proficiency to see whether different decoding challenges are encountered by learners at different stages of L2 development.

These gaps, along with those discussed in section (2.1.2.2.6.5), will be addressed in the present thesis.

2.3.The present study: Aims and research questions

The present study aimed to investigate decoding among Saudi (Arabic L1) ESL students, and the contribution of their second langauge decoding proficiency to their second langauge reading comprehension. Particularly, it aimed to:

(1) explore the L2 decoding performance of Saudi (Arabic L1) foundation year ESL students. Particularly, to explore how well they are able to perform on measures of L2 English decoding, and the nature of their decoding realisations, experiences, and processes.

(2) investigate the contribution of these learners' second language decoding proficiency to their second language reading comprehension, relative to the contributions of L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension.

In light of the above aims, the research questions that this study set out to answer are:

1. How well do Saudi foundation year ESL students (CEFR: A1, A2, B1) perform on measures of English real word and pseudoword decoding?
2. What is the nature of these Saudi ESL students' decoding realisations, experiences, and processes?
3. What is the contribution of these students' L2 decoding proficiency to L2 reading comprehension, relative to the contributions of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension?

3. Methodology

This chapter is divided into seven sections, the first of which describes the research design of the present study. In the second section, a description of the sampling procedure (including a description of the sampling frame, size, and strategy) is provided. The third section outlines the study's overall procedure. In the fourth section, a detailed description of the quantitative and qualitative data collection methods is presented. The analytical techniques adopted to answer each of the research questions are then discussed in the fifth section. A discussion of the piloting phase is presented in the sixth section, followed by ethical considerations in the seventh and final section.

3.1.Research Design

A mixed-method design was adopted in the present study, best described as a QUAN-qual design, with quantitative data used as a primary source, and the qualitative component aiming to explain and provide greater depth to the quantitative results. The rationale for selecting this design is discussed below:

The first part of the study aimed to explore ESL decoding among Saudi foundation year ESL students. In particular, the sub-objectives were:

(1) to explore how well they are able to perform on measures of L2 English decoding; and

(2) to explore the nature of their decoding realisations, experiences, and processes.

To understand whether or not these learners encounter difficulties with decoding the second language, it was deemed necessary to quantitatively measure learners' decoding proficiency from the outset. Calculating a percentage of success on a comprehensive and widely used measure of English decoding allowed the researcher to obtain an understanding of the L2 decoding proficiency of learners at each CEFR level (A1, A2, and B1). Additionally, given that the decoding measures have been standardised for L1 English readers, it was also possible to provide additional information regarding the decoding abilities of these A1, A2, and B1 learners (for example, the L1 grade equivalents of their scores), thus adding further detail to the description of the decoding abilities of L2 learners. However, decoding test scores offer little insight into the *nature* of any potential decoding difficulties, nor to the nature of the decoding processes that students engage in, which might have contributed to the observed decoding realisations/scores. A qualitative approach to enquiry thus allowed the researcher to understand the nature of learners' L2 decoding difficulties which, in the present study, was

achieved by means of:

(a) an error analysis of their L2 decoding realisations (outputs), using audio-recorded decoding data.

(b) a follow up interview investigating learners' decoding experiences with L2 English decoding.

(c) an examination of learners' L2 decoding processes, which was achieved by means of retrospective verbal reports (think-aloud protocols).

In so doing, the researcher achieved an understanding of the nature of students' L2 decoding realisations, experiences, and processes, thus providing greater depth to the obtained decoding proficiency scores.

The second part of the study aimed to examine the contribution of L2 decoding proficiency to L2 reading comprehension among these low-proficiency ESL readers. Herein, the researcher had three primary objectives and a fourth secondary one. The first three objectives pertained to the examination of the following:

(1) whether decoding makes a significant contribution to L2 reading comprehension before and after the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability are partialled out;

(2) whether the contribution of L2 decoding proficiency to L2 reading comprehension, if any, varies as a function of the linguistic proficiency of the learner, and if yes, in what ways.

(3) whether measures of L2 real word and L2 pseudoword decoding make independent contributions to L2 reading comprehension;

With a secondary objective to:

(4) provide triangulation, and explanation, for the statistical contribution (or lack thereof) of L2 decoding proficiency to L2 reading comprehension performance.

A quantitative correlational methodology addressed the first three objectives. This was achieved by assessing the performance of Saudi foundation year ESL students on six measures: L2 real word decoding, L2 pseudoword decoding, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension ability. The first five measures constituted the independent variables, and the last measure constituted the dependent variable. Relevant multiple regression analyses were then carried out to address

these three objectives. Further, a qualitative component, in the form of interviews comprising task based discussions, was considered a suitable complementary tool to validate the quantitative method of investigation, explain relations (Cohen, Manion, & Morrison, 2011), and also to further understand the factors that might account for any unexplained variance in the regression model tested.

To sum up, a diagrammatic representation of the overall design is provided below (Figure 3).

L2 decoding in Saudi foundation year ESL students

<table>
<tr><td colspan="2">RQ1: How well do Saudi (Arabic L1) ESL students at a range of L2 proficiency levels (CEFR: A1, A2, B1) perform on measures of English decoding?</td><td colspan="6">RQ2: What is the nature of the decoding realisations, experiences, and processes of Saudi foundation year ESL students?</td></tr>
<tr><td colspan="2"></td><td colspan="2">➢ What is the nature of the decoding realisations of Saudi foundation year ESL students?</td><td colspan="2">➢ What is the nature of the decoding experiences of Saudi foundation year ESL students?</td><td colspan="2">➢ What is the nature of the decoding processes of Saudi foundation year ESL students?</td></tr>
<tr><td>Data collection method</td><td>Data analysis plan</td><td>Data collection method</td><td>Data analysis plan</td><td>Data collection method</td><td>Data analysis plan</td><td>Data collection method</td><td>Data analysis plan</td></tr>
<tr><td>Decoding proficiency tests:
• Real word decoding test
• Pseudoword decoding test
N=212 (following outlier analysis)
(A1=71, A2= 89= B1= 52)</td><td>• means and mean percentage of success were calculated for students at each level of proficiency. Also, the L1 grade-equivalents of these scores were obtained from TOWRE's examiner manual.</td><td>Students' performance on the decoding measures was audio recorded</td><td>• The recording data of 60 participants (A1=20, A2= 20, B1=20) were phonologically transcribed.
• The decoding difficulties of the whole group (N= 60) were coded, resulting in several error categories.
• Between-group qualitative differences: Data was checked to see whether learners at different levels of proficiency had similar/different categories of error.
• Between-group quantitative differences: A series of ANOVAS were conducted to see whether learners at different levels of proficiency had similar/different error rates in each of the emerging categories of error.</td><td>Semi-structured interviews comprising task-based discussions about students' experiences with English decoding (N=12)</td><td>Thematic analysis</td><td>Retrospective verbal reports (think-alouds) (N= 15)</td><td>Thematic analysis</td></tr>
</table>

Figure 2. Study design (part 1)

The contribution of second language decoding proficiency to second language reading comprehension

RQ3. What is the contribution of second language decoding proficiency to second language reading comprehension?

Issues covered

▶ The relationships between overall L2 decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge, L2 reading comprehension ability, and L2 reading comprehension ability	▶ The contribution of overall L2 decoding proficiency to L2 reading comprehension *before* controlling for the contributions of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability	▶ The contribution of overall L2 decoding proficiency to L2 reading comprehension *after* controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability	▶ The contribution of overall L2 decoding proficiency as a function of the proficiency level of the learner	▶ The independent contributions of L2 real word decoding proficiency and L2 pseudoword decoding proficiency to L2 reading comprehension	▶ The contribution of L2 decoding proficiency to L2 reading comprehension: qualitative insights from the experiences of second language readers.

Data collection method	**Data collection method**
The following six measures were used: L2 real word decoding proficiency (N=212), L2 pseudoword decoding proficiency (N=212), L2 vocabulary knowledge (N=212), L2 grammar knowledge (N=212), L1 reading comprehension ability (N=212), and L2 reading comprehension (N=212).	Interviews (N=18) including task-based discussions about students' experiences with the issues that interfere with L2 reading comprehension
Data analysis plan	**Data analysis plan**
A series of correlational and regression analyses	Thematic analysis

Figure 3. Study design (part 2)

3.2.Sampling

3.2.1. Population

The population of interest for this research was Saudi foundation year ESL students. This population was selected for three reasons: First, English is gradually becoming the language of instruction in Saudi HE institutions. To address language-related challenges, universities started to deliver compulsory English courses for foundation year students (Ebad, 2014). Given the recency of this phenomenon, very little ESL research has thus far been conducted in this context. Particularly, no research has previously explored L2 decoding (abilities, realisations, experiences, or processes) among the population of foundation year ESL students, nor the contribution of their L2 decoding proficiency to their L2 reading comprehension performance. Second, it was anticipated that foundation year students would exhibit a good range of ESL abilities. This in turn would produce a set of test results that do not suffer from ceiling/floor effects which would allow carrying out the necessary statistical analyses. Thirdly, the researcher's affiliation with a Higher Education (HE) institution in Saudi Arabia made it easier to gain access to this year group.

3.2.2. Selection of the Higher Education Institution

The sampling site was one higher education institution in the western region of Saudi Arabia. At this university, as in most Saudi HE institutions, foundation year students take ESL as a compulsory subject to assist them in their academic studies (HEI, 2016). Upon their university admission, and before the foundation year commences, students take the Oxford Online Placement Test in proctored conditions, to assess their level of English proficiency (i.e., A1/beginner, A2/elementary, B1/below intermediate, or B1+/intermediate). Students demonstrating a level of proficiency beyond the intermediate level (by providing an IBT TOEFL score of 57 or above, or IELTS band score of 4.5 or above) are exempt from the English course (ibid).

It was firstly the need felt by the funding body (the HE institution) to understand the reading abilities of these students that led to the selection of this site. This site also offers several advantages: (1) there are around 15,000 foundation year students, a population which would allow the collection of a large dataset; (2) students take 18 hours of English classes a week (3-4 hours per day), which would facilitate group testing; (3) the homogeneous ESL curriculum in the HE institution is expected to keep the effects of curriculum/teaching methods at a minimum; (4) the researcher's previous working experience at the HEI has allowed her to establish contacts with authority figures which facilitated access and data collection.

At the outset, this purposive sampling may appear to limit generalization as foundation year students at other Saudi HE institutions may, among other things, (a) exhibit different ESL abilities, and/or (b) may be exposed to differential curriculum/teaching methods (for example, whether they receive/do not receive any decoding instruction). However, these two issues were less of a cause of concern in the present study. With regard to the first issue, it is worth mentioning that foundation year students in this HE institution, and in most HE institutions in Saudi Arabia, take an English placement test prior to university admission. The availability of this information on students' language proficiency makes it possible to tentatively generalize findings to students at other HE institutions *who exhibit similar levels of English proficiency*, i.e., low proficiency (A1, A2, and B1) ESL students. With regard to the second issue, the researcher had extensively reviewed relevant literature (e.g, see Alsubaie, 2014), examined various ESL curriculums, and engaged in conversations with various ESL teachers. These observations suggest that L2 decoding instruction is largely absent not only in the Saudi HE context, but also in the broad L2 pedagogical context more generally. Given the absence of decoding instruction, there was little reason to assume that students in this context would not be representative of proficiency-matched students at other Saud higher education institutions. That being said, however, there might be other variables, presently unknown to the researcher, which could differentiate the present study participants from students at other higher education institutions and which could have influenced the results of this study. Thus, care still needs to be exercised when generalizing the conclusions drawn from this study to the wider population of Saudi foundation year ESL students.

3.2.3. Participants' selection criteria

Having selected the site of data collection, inclusion/exclusion criteria were applied to all foundation year ESL students in this site. The criteria that participants had to meet were (a) Saudi Arabian nationals; (b) aged over 18; (c) learners who have not received instruction in reading a language other than Arabic and English; (d) absence of diagnosed clinical reading condition (for example, dyslexia); (e) having taken the language proficiency test, as students who miss taking the test are automatically classified as A1 level students.

Furthermore, after piloting, two additional criteria were added:

(f) Low proficiency ESL learners only (i.e., those classified as A1, A2, and B1). This criterion was added because, initially, the researcher planned to additionally include B1+ learners. However, it was discovered that a B1+ classification does not represent a "plus level" (i.e., the level in between B1 and B2) as the researcher had initially assumed. Rather, the classification includes learners who took the placement test and

achieved a score higher than the maximum possible score in the B1 level range, *regardless of* where that score stands (i.e., B2, C1, or C2). As such, B1+ students may exhibit a wide range of abilities that do not truly match the proficiency level they were assigned (i.e., intermediate).

(g) Female-only students. This criterion was applied because male and female foundation year ESL students' study in separate classes and the researcher was not allowed access to male only classes.

3.2.4. Sample size

The researcher was aiming at a sufficient sample size to support the core quantitative analysis in the present study, i.e., multiple regression. Following the formula for multiple regressions given by Tabachnick and Fidell (2014, p.123): $N > 50+8m$ where m equals the number of independent variables, a sample size of 90 was identified as the minimum size for carrying out the regression analyses needed. However, given that a large sample size could boost the power of the statistical tests performed, the researcher decided to exceed this minimum threshold and collect as many cases as the time resources allowed.

A total of 216 students completed the quantitative phase of the project (n=212 following outlier analysis). All of these foundation year students constitute low-proficiency ESL learners (the institute classified 71 of these students as beginner/A1 level students, 89 as elementary/A2 level students, and 52 as pre-intermediate/B1 level students). Of these students, 78 expressed interest in participating in the qualitative phase of the project. Of these, a subsample of 12 students participated in the interview into learners' experiences with English decoding (interview A) (see section 3.2.5 for sampling strategy). The institute classified 4 of these students as beginner/A1 level students, 4 as elementary/A2 level students, and 4 as pre-intermediate/B1 level students. Furthermore, a subsample of 18 students participated in the interview into readers' experiences of the issues that interfere with their L2 reading comprehension (interview B). The institute classified 6 of these students as beginner/A1 level students, 5 as elementary/A2 level students, and 7 as pre-intermediate/B1 level students. A subsample of 15 students participated in the think-aloud task. The institute classified 5 of these students as beginner/A1 level students, 5 as elementary/A2 level students, and 5 as pre-intermediate/B1 level students).

However, one of the qualitative phase participants did not take part in the quantitative phase but expressed interest in participating in the qualitative phase (interview B). Following a discussion with the classroom teacher, there was no reason to consider the student atypical to other students in the class in any way. Similarly, two of the participants who participated in

interview (A) did not take part in the quantitative phase but expressed interest in the qualitative phase. Because this sampling error was noticed in the analysis stage, it was not possible to check with the classroom teacher if the two students were atypical in anyway. However, as these students were placed in their classes based on their performance on the placement test, there was no reason to consider them different from other ESL students in the class.

3.2.5. Sampling strategy

The sampling strategy used in the quantitative phase of this study was random cluster sampling. The total population is divided into classes each comprising approximately 35 to 48 ESL students. Following teachers' expression of interest in the researchers' visit, nine ESL classes (clusters) were randomly selected to be visited. Cluster sampling is suitable when the characteristics of the clusters (classrooms) are similar (Cohen et al., 2011) – as is the case in the present research, (i.e., nationality, L1 background, and curriculum). It was also used because foundation year ESL students have roughly a full daily schedule of classes (from 8am to 4pm) and it was deemed very difficult to recruit participants outside of their ESL class time. Thus, testing needed to occur during ESL class time with intact classes. Cluster sampling was also suitable because students have ESL classes every day and these classes last for three to four hours ("morning classes" last for three hours, from 8am to 11am, and "afternoon classes" last for four hours from 11am to 4pm, with a lunch break in the latter from 12 noon to 1pm). The researcher also discussed her data collection plans with an experienced colleague at ELI who expressed that students were likely to participate if they knew that testing would take place during their ESL class time rather than outside it.

Howeve, it is worth noting that the researcher decided to not visit a class in certain cases, for example, (a) where the classroom-visit times suggested by the teachers were scheduled in the last teaching hour: that is, 10am to 11am for morning classes or 3pm to 4pm for afternoon classes. This is because, from the researcher's own teaching experience in this same site, students often feel fatigued in this last hour of the class, which in turn might affect their test performance; (b) where the classroom visit times suggested by the teachers were scheduled on a date judged to be too late by the researcher, as it might lengthen the data collection phase; (c) where the classroom teacher speculated that students in her class might not be interested in participating; (d) if the classroom teacher was behind in the pacing guide of the curriculum. This was done to minimize any risk of disruption to students' learning; and also (e) where the classroom-visit times suggested by the teachers preceded an English exam, and where rescheduling was not possible, as students might prefer to revise the test content with their ESL

teacher during that time, or might be more anxious/less focused, which might affect their performance on the tests.

Toward the end of the quantitative phase, the researcher visited the classes again to introduce the qualitative part of the project[7]. Random sampling was then applied for the selection of the qualitative phase participants. The sampling strategy, however, was not completely random because of the following: (a) where the researcher could not reach a participant to arrange the interview time (e.g., where a participant's phone was off or where they did not answer their phone), another participant was randomly selected; (b) where a participant suggested an interview time that was judged by the researcher to be too late or otherwise unsuitable, another student was selected.

3.3.Overall Procedure

The researcher held a meeting with the ELI Chief Coordinator and Head of Educational Affairs to introduce the project and ask them to inform the ESL teachers of the researcher's visit. Also, the researcher visited ESL teachers' offices and introduced the project herself to the teachers. Interested teachers were then asked to introduce the project verbally to the students and to inform them of the researcher's visit. Approximately a week later, the researcher visited the classes to introduce the project, outline the eligibility criteria and give out the information and consent sheets which were collected the following day. The ESL teachers were then contacted to arrange the dates and times of the two group testing sessions. In the first testing session, the L2 reading comprehension and L2 vocabulary knowledge measures were group-administered by the researcher during ESL class time and this session lasted for approximately an hour and 10 minutes. In the second testing session, the L2 grammar knowledge and L1 reading comprehension measures were group-administered also by the researcher and during ESL class time. This session took place within a maximum of one week from the date of the first testing session and lasted for approximately 50 minutes. After the second testing session, the researcher contacted the participants over the phone to arrange the date and time for the individual decoding testing session. This third and final session was scheduled outside of ESL class hours and often during students' break times or before the beginning of their first lecture (i.e., from 07:00-08:00). In this third session, the researcher verified the eligibility of the participating students by administering a brief background questionnaire. She then administered the two decoding measures (real word decoding and

[7] At the time the researcher was inviting students to take part in the qualitative phase, some students had stopped attending ESL classes to prepare for exams in other subjects (pre-exam week in the university). Therefore, not all of the 216 students were invited to participate in the qualitative phase of the project.

pseudoword decoding). This session took place within a maximum of one week from the date of the second testing session and lasted for approximately 10 minutes. Toward the end of the quantitative phase, the researcher visited the classes again to introduce the qualitative part of the project and administer the information and consent sheets. Students who expressed interest in participating in the interviews and retrospective verbal reports were later contacted over the phone to arrange the interview date and time. An outline of all of the data collection sessions is summarized in Table 2 below.[8]

[8] This procedure was followed unless otherwise was deemed suitable and/or necessary, e.g., where the classroom teacher was present in the class during the group testing sessions, the teacher often offered to help the researcher with administering the tests and the researcher allowed her to do so as it speeded the process of handing out the tests. Also, on rare occasions, students had to reschedule the date of the decoding testing session and the rescheduled date was only slightly outside the one-week window between the second and third sessions.

Table 2. Outline of the data collection sessions

<table>
<tr><td colspan="3">Session: One (233 participants)</td></tr>
<tr><td colspan="3">Method of administration: group testing</td></tr>
<tr><td colspan="3">Duration: 1 hour 10 minutes</td></tr>
<tr><td colspan="3">1. L2 Reading comprehension measure (45 mins)</td></tr>
<tr><td colspan="3">2. L2 Vocabulary knowledge measure (25 mins)</td></tr>
<tr><td colspan="3">Session: Two (227 participants)</td></tr>
<tr><td colspan="3">Method of administration: group testing</td></tr>
<tr><td colspan="3">Duration: 50 minutes</td></tr>
<tr><td colspan="3">3. L2 Grammar knowledge measure (25 mins)</td></tr>
<tr><td colspan="3">4. L1 Reading comprehension measure (25 mins)</td></tr>
<tr><td colspan="3">Session: Three (216 participants)</td></tr>
<tr><td colspan="3">Method of administration: individual</td></tr>
<tr><td colspan="3">Duration: 10 mins</td></tr>
<tr><td colspan="3">5. Decoding measures:
5.1. Real word decoding measure (2.5 mins)
5.2. Pseudoword decoding measure (2.5 mins)
6. Background questionnaire (5 mins)</td></tr>
<tr><td colspan="3">Participants expressing interest in participating in the qualitative phase were then invited to one of the following sessions (depending on their preference)</td></tr>
<tr><td>Session: Four (15 participants)</td><td>Session: Five (12 participants)</td><td>Session: Six (18 participants)</td></tr>
<tr><td>Setting: individual</td><td>Setting: individual</td><td>Setting: individual</td></tr>
<tr><td>Duration: 15 mins</td><td>Duration: 8 mins</td><td>Duration: 12 mins</td></tr>
<tr><td>Think-aloud protocols

<u>Objective:</u> to investigate the decoding processes of Saudi foundation year ESL students.</td><td>Interview A (comprising task-based discussion)
<u>Objective:</u> to investigate the decoding experiences of Saudi foundation year ESL students.</td><td>Interview B (comprising task-based discussions)
<u>Objective:</u> to investigate students' experiences of the issues that interfere with their reading comprehesnion.</td></tr>
</table>

3.4.Data collection methods

3.4.1. Quantitative data collection methods

3.4.1.1.Measures

3.4.1.1.1. L2 decoding proficiency

A reading aloud task was used to test the participants' L2 proficiency. This is the standard decoding task in first language and second language reading research (García & Cain, 2014; Jeon & Yamashita, 2014). As decoding proficiency in the present study includes the ability to obtain phonological codes from graphical codes in the potential presence and absence of prior phonological knowledge (i.e., when a participant can/cannot rely on prior phonological knowledge of the items), both the Sight Word Efficiency and the Phonemic Decoding Efficiency subtests of the Test of Word Reading Efficiency (TOWRE) were used as a measure of *overall L2 decoding proficiency* (Torgesen, Rashotte, & Wagner, 2012). This approach of including both real words and pseudowords is thought to provide the most comprehensive estimate of readers' decoding proficiency (Adlof, Catts, & Little, 2006; García & Cain, 2014).

TOWRE was selected as a measure of decoding proficiency for several reasons: First, unlike other decoding measures (e.g., Woodcock, 1987), it assesses readers' ability to decode words within a time limit. Thus, it provides a measure not only of readers' decoding accuracy but also of their decoding speed. Second, the publisher provides high reliability and validity evidence for TOWRE. For example, the test-retest reliability values ranged from .86 and .90 for the real word and pseudoword subtests, respectively, and concurrent validity values ranged from .89 to .94 when TOWRE was compared with Woodcock's test (Woodcock, 1987); (see Torgesen, Rashotte, & Wagner, 2012 for detailed validity and reliability evidence). Third, TOWRE is a standardized test which offers two advantages: (a) standardized tests, being widely used, facilitate rigorous comparability of research findings (for example, whether the decoding abilities of A1, A2, and B1 students in this study are similar, or diverge from, those of previous/future studies using the same measures; whether the contribution of decoding proficiency to reading comprehension in this population is similar to, or diverges from, that of previous/future studies using the same measures); (b) as the test has been standardized among a large population of L1 readers (grades 1-12+, ages 6-24), TOWRE would additionally allow us to analyze the decoding abilities of L2 learners from a comparative perspective. For example, what the equivalent L1 grade levels that learners at A1, A2, and B1 levels of proficiency perform at. This could enhance the field's understanding of the decoding abilities of L2 learners which, to date, remains vague and lacking in specificity. Fourth, TOWRE can

be administered in a relatively short time which was deemed practical given the somewhat large sample size.

However, one issue with TOWRE (and indeed with other standard and widely-recognized decoding measures, such as Woodcock (1987)) is that allowance is made for the real word subtest to include items which may be completely unfamiliar to the test takers. For example, the test might include words which the L1 reader, and by extension the L2 reader, might not have encountered before, be it their oral and/or their written form. This is despite the fact that a real word must be encountered at least "once" for it to be decoded holistically on subsequent encounters (Hoover and Gough, 1990). The developers of TOWRE, however, do acknowledge this limitation in their test. As they put it:

> "Although many of the words on [the real word] subtest can be identified through phonemic decoding processes, as individuals move to higher levels of reading development, they cannot obtain an average or above average score on this subtest without being able to recognize most of the words by sight (Torgesen, Rashotte, & Wagner, 2012, p. 64)".

Herein, the researcher was left to decide whether to continue to make this allowance (i.e., to retain the original format of the real word subtest) or to revise the subtest in such a way that it strictly includes items that the test takers are familiar with (for example, by identifying items which they are likely to have never encountered before and removing these from the test). A decision was made to retain the original format of the test and to acknowledge the possibility that the test may include items which might be unfamiliar to some participants. This was done for four reasons. First, one concern with devising a test comprising very familiar items was that it could lead to minimal variability in the real word decoding test scores, thus leading to a ceiling effect (a problem often reported in the literature and which might result in the inability to detect a relationship between decoding and reading comprehension). Second, during piloting, almost all the items (particularly, those read within the 45 second testing period permitted by the test instructions) were reported to have been encountered previously (regardless of whether or not their meanings were familiar), with the average percentage of the unfamiliar items being 8.4%[9]. This figure was judged to be small and did not cause serious concerns. Third, and as previously mentioned, in their construct validation study, TOWRE's developers reported that the real word and pseudoword tests loaded onto two different factors *despite* the allowance made for the real word test to include unfamiliar items (Torgesen,

[9] or 4% using a lenient method, e.g., where comments along the lines of "I think I have seen it but I'm not sure" were not counted as unfamiliar.

Rashotte, & Wagner, 1999) which suggests that the inclusion of the unfamiliar items did not impact (at least significantly) the validity of the measure. Fourth, it is worth mentioning that, as a follow-up to the findings obtained in the present study, the researcher calculated the word frequency bands that the items in the real word subset of TOWRE belong to (see Appendix F). The findings showed that the majority of the test items (65%) belonged to the most frequent 1000 word frequency band and that 98% of the first and second quartiles of the test (i.e., the items presented in the first half of the test) belonged to the 1000 word frequency band. In fact, approximately a quarter of the items there were even function words such as <is>, <me>, <he>, <to>, <no>, <yes>. Furthermore, and most importantly, on average 94% of the items read by participants in the present study belonged to the 1000-word frequency band. As Saudi foundation year ESL students would have learned English for a minimum of 8 years in state schools (12 years in private schools), there was no reason to question their prior encounter with these highly frequent words. (Further relevant support can be found in the think-aloud findings reported in Results chapter I, where approximately 81% of the items decoded "automatically" were real words rather than pseudowords, which were by definition unfamiliar to all participants).

The two decoding subtests are discussed below.

3.4.1.1.1.1. L2 pseudoword decoding proficiency

The pseudoword subtest contained 66 items. The student is instructed to read aloud the items as fast as they can within 45 seconds (using a stopwatch). The test was administered in a one-to-one interview format in one of the university rooms and as per the publisher's administration guidelines. A trial session containing 8 items was administered to each participant before the start of the main testing session.

To judge decoding accuracy, the examiner's manual provides real words (e.g., <late>) that rhyme with each of the target pseudowords (e.g., <pate>) and items were judged as correct/incorrect as per this marking scheme. However, as in previous L2 studies (e.g., Hamada and Koda, 2008), some allowance was made for foreign accent when scoring the test. It should be noted here that it was challenging to decide which error would be counted as a foreign accent (and thus scored correctly) and which as an error (and scored incorrectly). This is because the researcher had concerns about making excessive allowance for foreign accent, especially in the case of Arabic ESL readers. Doing so would mean that a decoding of, say, <left> as /lift/, <van> as </fæn/, and <pack> as <bæk> could be considered correct, even though these realisations completely alter the identity, and meaning, of the word. This is particularly

problematic in research such as the present one where the impact of decoding on reading comprehension is of interest. This is because decoding errors of this nature could potentially lead to comprehension breakdowns. For instance, it could contribute to inaccurate meaning access or, even if access was not disturbed, it could potentially disrupt subsequent comprehension monitoring as the realisations of <lift> and <left> would not be distinguishable in the phonological loop of the reader (as both are realized as /lift/). In fact, it is interesting and relevant to mention here that in a study where Arabic participants were asked to translate written English words to Arabic, one of the error themes that emerged was misinterpretations caused by phoneme confusions, e.g., confusing the meanings of <pillow> and <blow>, <invent> and <infant>, etc. (note that Arabic lacks the phonemes /p/ and /v/ and readers often substitute these with /b/ and /f/, respectively) (Alhazmi 1993, cited in Emrey (2005)).

Therefore, to decide which errors should be considered as 'decoding errors' per se (reflecting incorrect GPC) and which were simply the result of foreign accent, intelligibility was used as the key criterion in the present study. Examples of foreign accent allowances include a reading of <r> as a rolled /r/ (e.g., when decoding the item <river>) and a reading of /ŋ/ as /ng/ (e.g., when decodeing the item <morning>).

Both inter-rater and intra-rater reliability were then examined for 5% of the cases (N=11). To judge intra-rater reliability, the researcher re-marked the tests of these cases. Then, each participant's score on each test item (N of items= 66, i.e., 726 individual items in total) was recorded in time 1 and 2. The data was then subjected to reliability analysis using Cohen's kappa. Kappa estimates were selected because they take into account the likelihood of chance agreement (unlike traditional percentage agreement methods). Correcting for chance agreement was a crucial consideration in the present study, as the categories were binary (correct vs. incorrect) and, as such, chance agreement is more likely to occur. Thus, kappa would allow for a stringent test for the levels of agreement at the two time points. An overall kappa was then calculated by averaging the individual kappa values. There was an almost perfect agreement between the researcher's scores at Time 1 and Time 2, k= 0.92.

To judge inter-rater reliability, a near native English and Arabic speaker – a PhD candidate in Phonetics and Phonology – was asked to find the incorrect decoding items also in 5% of the cases (N=11). Then, each participant's score on each test item (N of items= 66, i.e., 726 individual items in total) was recorded. Again, the data was then subjected to reliability analysis using Cohen's kappa. An overall kappa was then calculated by averaging the

individual kappa values. There was a substantial agreement between the researcher and inter-rater's accuracy judgements, average k= 0.79.[10]

However, because there was concern that the figures may be overestimated by the presence of "automatic agreements" (i.e., caused by missing items), the researcher re-ran the analysis using the attested realisations only and found that there was an almost perfect agreement in the intra-rater reliability figures k= 0.89, and substantial agreement in the inter-rater reliability figures k= 0.69.

3.4.1.1.1.2. L2 real word decoding proficiency

This subtest contained 108 items. The subtest was administered in a one-to-one interview format and as per the publisher's administration guidelines. The student was instructed to read aloud the items as fast as they can within 45 seconds (using a stopwatch). A practice session containing 8 items was administered to each participant before the start of the main testing session. To judge decoding accuracy, the researcher marked the items as correct/incorrect as per the marking scheme provided in the Examiner's Manual. However, as in the pseudoword test, some allowance was made for foreign accent.

Both inter-rater and intra-rater reliability were then examined for 5% of the cases (N=11). To judge intra-rater reliability, the researcher re-marked the tests of these 11 cases. Each participant's score on each test item (N of items= 108, i.e., 1188 individual items in total) was recorded at times 1 and 2. The data from each case was then subjected to reliability analysis using Cohen's kappa. An overall kappa was then calculated by averaging the individual kappa values. There was an almost perfect agreement between the researcher's scores at Time 1 and Time 2, average k= 0.94, p<.000. To judge inter-rater reliability, the above mentioned inter-rater was asked to find the incorrect items also in 5% of the cases (N=11). Then, each participant's score on each test item was recorded. The data was then subjected to reliability analysis using Cohen's kappa. An overall kappa was then calculated by averaging the individual kappa values. There was an almost perfect agreement between the researcher's and inter-rater's judgements, average k= 0.91.

However, again because there was concern that the figures may be overestimated by the presence of "automatic agreements" caused by missing items, the researcher re-ran the analysis using the attested realisations only and found that there was an almost perfect

[10] According to Cohen, kappa values of ≤ 0 indicate no agreement, 0.01–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41– 0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as almost perfect agreement (Cohen, 1960).

agreement in the intra-rater reliability figures, k= 0.88, and substantial agreement in the inter-rater reliability figures k= 0.74.

3.4.1.1.2. L2 reading comprehension

The reading comprehension subtest of institutional tests was used to assess learners' L2 reading comprehension. These are multiple-choice tests which were developed by a panel of experts in ESL assessment and which are based on CEFR-linked test specifications (HEI, 2016). The test items are designed to tap both literal and inferential comprehension (i.e., information explicitly stated in the text and information that requires readers to link text information). One passage was selected from the subtest of each proficiency level (A1, A2, B1) resulting in a test comprising three passages with six items each (18 items in total). However, as – at the time of data collection – B1 students might potentially find the passages at their level easily comprehensible, which could result in a ceiling effect, the researcher added one passage from the B1+ subtest (6 items). This resulted in a 24-item test divided between 4 passages progressing in difficulty (Appendix A). The passages are short to medium in length with the shortest containing 158 words and the longest containing 385 words.

The test was group-administered by the researcher during the first part of session 1. Instructions were provided orally and in the participants' L1 (Arabic). Participants were asked to read each passage (starting from passage one), answer the items of the passage, and then to move to the next passage. They were reminded to avoid blind guessing if they had no idea what the answer might be. The test took approximately 45 minutes to administer. Scoring was dichotomous with each correct answer receiving one point and each incorrect answer receiving zero points.

To assess the reliability of the L2 reading comprehension measure, as per Jeon (2011), 20% of the cases were randomly selected (N= 43). Then, each participant's score on each item of the test was recorded (N of items= 24). The data was then subjected to reliability analysis using Cronbach alpha. The measure appeared to have good internal consistency, as determined by a Cronbach alpha of .84 (Kline, 2000).

Institutional reading comprehension tests were selected for three reasons: First, the researcher's work experience in this HEI has allowed her to examine first-hand the suitability of their institutional tests (i.e., having previously reviewed the tests, administered them to her students, and having been in a position to receive any test-related feedback from students). As this experience was very positive, the researcher considered these as good candidates for the present study. Second, there is some indirect evidence for the predictive validity of institutional

tests. The HEI (2016) seeks to ensure that students who successfully complete the English course are at intermediate proficiency level by the end of the foundation year. Therefore, the language institute conducted a study in which 150 students (who completed the foundation year English course) were entered for the IELTS exam. The students scored an average of 4.5 on IELTS, suggesting that they were at intermediate proficiency level (ibid). Given that these students had to pass institutional tests to complete the English course, this finding could also be taken as evidence for the predictive validity of the institutional tests. Third, prior to selecting institutional tests, the researcher reviewed other candidate measures of reading comprehension, particularly TOEFL and standardized L1 reading comprehension measures, and these were judged to be problematic for many reasons. For example, upon reviewing TOEFL (the researcher obtained a copy of the test after preparing a proposal to ETS, Appendix H), the researcher and two academic colleagues at the HEI judged it to be too difficult for the participants, particularly those at A1 and A2 levels (see piloting phase). Its selection would thus not contribute to differentiating between participants in their reading comprehension skill, potentially resulting in a floor effect and an absence of correlations with all independent variables. In fact, in one study, Tannenbaum and Wylie (2008) asked language experts, representing 16 countries, to link CEFR levels to TOEFL scores. The panelists were unable to establish any linkages between the test and the A1 and A2 levels, considering the test inaccessible to learners at these levels, and only slightly accessible to B1 learners. The wide use of TOEFL in L2 reading research is, however, understandable given that most L2 reading research has been conducted with learners at the higher end of the proficiency scale (i.e. intermediate and above), often those residing in English-speaking countries where much research is conducted. As for L1 standardised reading measures, two major issues arose when these measures were carefully examined. First, the majority required individual administration (except the Gray Silent Reading Test (Wiederholt & Blalock, 2000), which can be administered in groups). Given the large number of participants needed in the current study, and their heavily-packed timetable of lectures, an individually administered reading test was deemed inappropriate. Second, and most importantly, since these measures are intended for use with L1 readers, the topics of low/beginner level passages were deemed unsuitable for adult users (see below a sample of a reading passage from the Gray Silent Reading Test (Wiederholt & Blalock, 2000)). [11]

[11] This is not to mention that various criticisms exist over these tests, e.g., those pertaining to their inclusion of passage independent items, as shown for the Nelson-Denny Reading Test by Coleman, Lindstrom, Nelson, Lindstrom, and Gregg (2010) and for the Gray Oral Reading Test (GORT) by Keenan and Betjemann (2006), or those pertaining to the unnatural format of the test (e.g., cloze) as in the Woodcock test (Woodcock, 19

Table 3. A sample passage from Gray Silent Reading Test (Form A)

The boy is not big.
But he helps Mother and Father.
He puts his toys up after he plays with them.
He tells a funny story if she begins to cry.
He sings to make her go to sleep.

3.4.1.1.3. L2 vocabulary knowledge

Vocabulary knowledge is multidimensional, but central to it is knowledge of word meaning (Koda, 2005). In research where reading comprehension is the dependent variable, vocabulary knowledge is often defined as the breadth of word knowledge, i.e., how many words a reader knows the meaning of (Jeon, 2011; Perfetti, 2010). Likewise, in the present study, vocabulary knowledge reflected the breadth of word knowledge and the Vocabulary Levels Test (VLT) was used for this purpose. The VLT offers a good measure given its high validity and reliability (See Schmitt, Schmitt, and Clapham (2001) for detailed validity and reliability evidence). Further, in the absence of standardized L2 vocabulary knowledge measures, the VLT has been described as the "nearest thing we have to a standard test in vocabulary" (Meara, 1996, p. 38). Also, unlike other vocabulary measures that need to be administered individually (for example, Peabody Picture Vocabulary Test (Dunn & Dunn, 1997); the VLT can be administered in group settings, making it more suitable for large sample sizes.

The test consists of 30 items in each of its five sections (with each section testing vocabulary knowledge of the most frequent 2000, 3000, 5000, 10,000 words, and academic vocabulary). Only the 2000 and 3000 sections were administered given the researcher's and two academic colleagues' judgement about the test's suitability to learners at the lower end of the proficiency scale (see piloting phase). The test (Appendix B) required participants to match words with their correct meanings. Below is a sample vocabulary task from the VLT:

1.	business		
2.	clock		
3.	horse	- - -	part of a house
4.	pencil	- - -	animal with four legs
5.	shoe	- - -	something used for writing
6.	wall		

Figure 4. A sample VLT task

However, given that the VLT is a written vocabulary test, i.e., participants read words before selecting their corresponding meanings, there was the possibility of decoding proficiency affecting performance on the test. To guard against this, the researcher initially decided to read aloud the test items to the participants to ensure their decoding proficiency did not affect their performance. However, when this was done in the piloting stage, it proved problematic because students did not proceed with the test items at exactly the same time. Therefore, following Zhang and Koda (2012), the researcher asked students to raise their hands when they encountered difficulties in reading any item. The researcher then provided the accurate reading (i.e., pronunciation) to that student individually. This interrelation of reading constructs is not easy to avoid, as Carr and Levy (1990) note in their seminal work on componential approaches to reading comprehension:

> It is literally impossible to devise a task that requires only one mental operation for its performance. Therefore, it will generally be the case that no task is a pure measure of a single component... Therefore, the goal must be to choose tasks whose performance reflects a particular component skill as strongly as possible and the influence of other processes as weakly as possible. If this criterion is met for all the hypothetical skills, then comparison among tasks in the battery will allow one to determine the relative importance of their contributions to [overall reading comprehension ability].

The VLT was group-administered by the researcher directly after the L2 reading comprehension measure (session 1). Instructions were provided orally and in the participants' L1 (Arabic). Participants were then asked to look at a sample task provided at the beginning of the test and to ask any questions. Afterwards, they were asked to answer the 2000 level section followed by the 3000 level section. Participants were reminded to avoid blind guessing if they had no idea what the answer might be. The test took approximately 25 minutes to administer. Scoring was dichotomous with each correct answer receiving one point and an incorrect answer receiving zero points.

To assess the reliability of the L2 vocabulary knowledge measure, as per Jeon (2011), 20% of the cases were randomly selected (N= 43). Then, each participant's score on each test item (N of items= 60) was recorded. The data was then subjected to reliability analysis using Cronbach alpha. The measure appeared to have an excellent internal consistency, as determined by a Cronbach alpha of .91 (Kline, 2000), a figure very similar to that reported in Schmitt et al's (2001) validation study.

3.4.1.1.4. L2 grammar knowledge

In this study, grammar knowledge is defined as morphosyntactic knowledge, specifically, "knowledge about a variety of morphosyntactic properties, i.e. tense, aspect, word order, subject–verb agreement, and articles" (Jeon & Yamashita, 2014, p. 165) which, as Jeon and Yamashita noted, is the most common definition in reading research. It is thought that it is this morphosyntactic information that provides ongoing instructions for the construction of text comprehension (Grabe, 2009).

The researcher reviewed popular existing grammar measures, such as Shiotsu's (2010), and judged them to be too difficult for the participants, particularly those classified as A1 and A2. For instance, learners at A1, A2, and B1 levels are introduced to the simple, progressive, and perfect aspects of the present and past tenses only. However, many of the items used in Shiotsu (2010) address the use of tenses at higher proficiency levels such as the future perfect which is normally introduced to B2-level learners. This is exemplified in the sample task below from Shiotsu (2010):

By the time this course finishes - - - - - - a lot about engineering.			
A. I will learn	B. I learn	C. I will have learnt	D. I have learnt

Figure 5. A sample grammar task from Shiotsu (2010)

Shiotsu's (2010) study involved participants from a very wide range of proficiency levels; nonetheless, the majority were English-majors who were at relatively high proficiency levels. Indeed, so far, the test has only been used in studies recruiting learners at the higher end of the proficiency scale (e.g., see Kremmel, Brunfaut, & Alderson, 2015), with no evidence of its suitability to low-proficiency readers currently existing. Thus, despite good validity and reliability evidence, there were serious concerns that the selection of this test would result in a floor effect and an absence of correlations that might otherwise have been present, if the test had been more accessible to all participants.

Therefore, the grammar subtest of institutional tests was used. The test items are derived from the grammar domains in three CEFR-linked course materials: *Cambridge English Unlimited (beginner, elementary, pre-intermediate)* (Doff & Robinson, 2013; Rea, 2010; Tilbury, Clementson, Hendra, & Rea, 2016). The items measure a wide variety of grammatical knowledge such as tense, aspect, word order, subject–verb agreement, articles, etc.

Furthermore, the content of the test was compared against that of another published grammar measure, Ellis's (2005), and a large degree of overlap was observed in the grammatical structure included.[12] The subtest of each proficiency level has 16 multiple-choice items. Seven items were selected from each subtest (proficiency level). When selecting the items, care was taken to ensure they did not address the same underlying knowledge, for example, where an item testing knowledge of the present simple was found in two subtests (A1-A2), only one was selected. This resulted in a test of 21 items (seven items from each subtest). However, as, at the time of data collection, B1 students may have mastered the grammar rules at their level, the researcher added 7 items from the B1+ level subtest and another 7 items from Shiotsu's (2010) test. This resulted in a 35 (multiple-choice) item-test (Appendix C).

The test was administered in groups by the researcher during the first half of the second session. Instructions were provided orally and in the participants' L1 (Arabic). Participants were reminded to avoid blind guessing if they had no idea what the answer was. The test took approximately 25 minutes to administer. Scoring was dichotomous with each correct answer receiving one point and an incorrect answer receiving zero points.

To assess the reliability of the L2 grammar knowledge measure, as per Jeon (2011), 20% of the cases were randomly selected (*N*= 43). Then, each participant's score on each test item (*N* of items = 35) was recorded. The data was then subjected to reliability analysis using Cronbach alpha. The measure appeared to have good internal consistency, as determined by a Cronbach alpha of .85 (Kline, 2000).

3.4.1.1.5. L1 reading comprehension

The Arabic reading comprehension test comprised of 5 reading passages with a total of 20 multiple-choice items. The test passages varied in length (120-518 words) and the items were designed to tap both literal and inferential comprehension. This test was adopted from a sample of the General Aptitude Test, a test developed by The National Centre for Assessment in Higher Education for university admissions in Saudi Arabia. The researcher was not allowed access to original exam papers but has been informed that sample (practice) materials go through the same rigorous development procedure as original exams (personal communication, 2016). Dimitrov and Shamrani (2014) reported an internal consistency of .77 for the Arabic reading comprehension subtest of the GAT.

[12] The grammar measure used in the present study includes all the grammatical structures covered in Ellis (2005) except four (dative alternation, ergative verbs, relative clauses, and plural s) which are, with the exception of plural s, often introduced to intermediate and advanced ESL students, and therefore, are not suitable to the present study participants.

The test (Appendix D) was group-administered by the researcher in the second half of the second session right after the students completed the L2 grammar knowledge test. Instructions were provided orally in Arabic. Participants were reminded to avoid guessing if they had no idea what the answer might be. The test took approximately 25 minutes to administer. Scoring was dichotomous with each correct answer receiving one point and an incorrect answer receiving zero points.

To assess the reliability of the L1 reading comprehension ability measure, as per Jeon (2011), 20% of the cases were randomly selected (*N*= 43). Then, each participant's score on each test item (*N* of items = 20) was recorded. The data was then subjected to reliability analysis using Cronbach alpha. The measure appeared to have an acceptable level of internal consistency, as determined by a Cronbach alpha of .68 (Kline, 2000).

3.4.1.1.6. Background questionnaire

A background questionnaire was prepared to be administered to the participants before the start of session 1 (though see below). The questionnaire did not intend to play any formal role in answering the research questions; however, it served two purposes:

Firstly, it aimed to verify the eligibility of the participants (e.g., nationality, age, proficiency level, absence of any clinical reading condition). Secondly, the data provided information on the other languages that students are familiar with, if any, and their proficiency level in those languages. This allowed an indication of the various phonemic systems that students might have access to, which aimed to inform and enhance the phonemic transcription process. Where languages other than those known by the researcher were found, the researcher sought to familiarise herself as much as possible with the phonemic inventories of those languages (e.g., by watching educational YouTube videos) to ensure the phonological transcription process was not biased by her lack of familiarity to certain sounds.

As noted above, the initial plan was to administer the questionnaire at the beginning of the first group testing session. However, upon visiting the data collection site, the researcher realised that much of the data generated from the questionnaire was self-evident. (For example, the HEI is a state university which require students to be of a Saudi nationality, the ESL proficiency level was already known from the class the students were enrolled to, the age criterion was logically met as the Saudi education system does not allow grade skipping, so all university level students would be over 18, and the ELI Student Support Unit has informed the researcher that no cases of dyslexia were reported). Additionally, during the invitation stage, the researcher had clearly outlined these eligibility criteria and students who did not meet these

criteria were not invited to participate. Therefore, there were concerns over the true usefulness of many of the questionnaire questions. Furthermore, it is worth noting that during piloting, some of the questionnaire questions triggered some time-consuming discussions, for example., "If I only know some words in language x should I consider that a third language?", "How can I judge my proficiency level in the third language?". As the researcher had to administer the L2 reading comprehension and L2 vocabulary knowledge measures in approximately one hour, it was deemed more appropriate to free the session time for these measures, particularly given that most of the questionnaire data was already available to the researcher.

Nonetheless, it was thought useful to collect information on students' exact ages, the languages that they are familiar with, if any, and their proficiency level in those languages. Therefore, a decision was made to collect this data before the beginning of session 3 (individual decoding session). This allowed for any discussions regarding students' knowledge of previous languages to take place.

3.4.1.2. Summary of the tests

Table 4 summarises all the measures used in the present study.

Table 4. Summary of measures

Test	Skill assessed	Approximate duration of administration	Method of administration	Reliability
TOWRE sight word reading efficiency (Form A)	Real word decoding	2.5 mins	individual	Cohen's kappa of .94 and .91 for inter- and intra- rater reliability, respectively
TOWRE phonemic decoding efficiency (Form A)	Pseudoword decoding	2.5 mins	individual	Cohen's kappa of .92 and .79 for inter- and intra- rater reliability, respectively
Institutional tests: Reading comprehension subtest	L2 reading comprehension	45 mins	group	Cronbach's alpha of .84.
The Vocabulary Levels Test	L2 vocabulary knowledge	25 mins	group	Cronbach's alpha of .91
General Aptitude Test: Arabic reading comprehension subtest	L1 (Arabic) reading comprehension	25 mins	group	Cronbach's alpha of .68.
Institutional tests: Grammar knowledge subtest	L2 grammar knowledge	25 mins	group	Cronbach's alpha of .85.

3.4.2. Qualitative data collection methods

3.4.2.1.Audio-recordings of learners' decoding outputs

One common method used to understand the nature of decoding difficulties is to audio-record students' decoding outputs. Audio-recording data allows for thorough examination and analysis of students' decoding errors, and hence difficulties (e.g., see Greenberg et al., 2002).

Thus, in the present study, students' performance on the two decoding measures was audio-recorded using an Apple iPhone voice recorder. 60 cases (20 per proficiency level) were then randomly selected for their decoding realisations to be transcribed using the International Phonetic Alphabet (IPA).

All realisations were phonemically transcribed by the researcher, an experienced phonemic transcriber who had previously completed a course in phonetics and phonology. However, competence in phonemic transcription was not the only crucial criterion here. In the field of

phonetics, there is growing awareness of the impact of transcribers' linguistic backgrounds on transcription validity. Research in this field suggests that the mother tongue of the transcriber (Cucchiarini, 1993), and even their dialect (Coussé, Gillis, Kloots, & Swerts, 2004), could affect their perceptions, and transcription, of sounds. For example, it is common for an English speaker to perceive the Arabic word /ħʌlə/ <hala> (meaning "dessert") as /hʌlə/ <hala> (meaning "hello/welcome") as the English phonemic inventory lacks the Arabic sound /ħ/. Given the researcher's familiarity with the L1 of these readers (Modern Standard Arabic and Saudi Arabic), there were no concerns over her ability to perceive the sounds of these phonemic systems.

When transcribing the decoding outputs, the following criteria were applied:

a) Where a decoding output was inaudible, that decoding output was excluded from transcription (for example, where a learner subvocalizes/mumbles the test item)
b) Where a phonemic sound was audible but not entirely clear (and thus difficult to categorise), the phoneme judged to be closest to the perceived sound was transcribed (for example, where a learner articulates a particular phoneme only partially as – during the course of its production – they seem to decide to shift to a different phoneme).
c) Outputs judged to be decoded correctly were excluded from transcription (because that would have been a redundant step: their decoding realisation was known anyway).
d) Where two or more decoding outputs were produced for a single item, only the last decoding output was transcribed.
e) As transcription was based on phonemic sounds (i.e., broad transcription), phonetic details were excluded from transcription. To illustrate, <pen> may be transcribed as /pen/ and [p^{h}en], the former being the broad/phonemic transcription, whereas the latter is the narrow/phonetic transcription, as it not only indicates that the word starts with the phoneme /p/ but also that the phoneme is aspirated as shown by the superscript [^{h}].
f) The schwa /ə/ and wedge /ʌ/ symbols were allowed to be used interchangeably. However, effort was made to use the former in unstressed syllables, for example, /ˈmʌ**ðə**/, and the latter in stressed ones, for example, /ˈ**mʌ**ðə/, and monosyllabic words, for example, /nʌts/.

Following preliminary pen and paper transcription, the transcriptions of 60 randomly selected cases (N=20 per level of proficiency) were digitised by the researcher to allow for relevant subsequent analyses.

3.4.2.1.1. Inter- and intra- transcriber reliability

The researcher's review of the literature on phonological transcription suggested that "transcriber reliability" has been the most widely used metric for transcription evaluation. As noted earlier, the phonemic transcription data was used to analyse learners' decoding realisations. Thus, it was crucial to ensure that this data was reliable. To this end, an intra- and inter- transcriber reliability study was conducted using a random sample of 5% of the cases (N=11).

3.4.2.1.1.1.Intra-transcriber reliability

To examine intra-transcriber reliability, the researcher re-transcribed the real word and pseudoword decoding outputs of these 11 cases. Percentage agreement was then calculated at the individual phoneme level, which is the level of analysis often studied in transcription reliability research (Cucchiarini, 1993). The agreement values for each decoding output lay on a scale ranging from 0% to 100% based on the number of phonemes transcribed identically. For example, where the researcher transcribes [men] as /mæn/ in time 1, but as /min/ in time 2, percent agreement for this item is 66.66% as two out of three phonemes were transcribed identically. It is worth mentioning that, using this analysis method, calculations were not always straightforward. This is because disagreements did not always pertain to differences in the phonemes transcribed, but also to the *number* of phonemes transcribed making it difficult to align the two transcriptions (e.g., 3 phonemes in time 1 but 4 phonemes in time 2). On such occasions, the longer transcription was used as the denominator. To illustrate, on one occasion, the researcher transcribed a learner's realisation of <lost> as /lɒts/ (4 phonemes), and the inter-transcriber as /lɒt/ (3 phonemes) (the inter-transcriber commented that the final /t/ was heavily aspirated giving the illusion of a /s/ sound). In cases like these, the number of the phonemes in the longer transcription (4 in the previous example) was used as the denominator. Using this example, the researchers would have agreed on the transcription of three out of four phonemes resulting in an item agreement level of 75%. Following calculating the percent agreement for each individual test item, the average percent agreement was calculated for each case (see Appendix L). Intra-transcriber reliability figures were extremely high, 98.43% and 96.16% for the real word and pseudoword decoding outputs, respectively. However, there was concern that these figures may be overestimated by the fact that correct items were automatically coded as

agreements. Therefore, the researcher re-analysed the data after excluding these items. In this analysis, intra-transcriber reliability figures were 93.50% and 92.37% for the real word and pseudoword decoding measures, respectively.

3.4.2.1.1.2.Inter-transcriber reliability

As previously mentioned, familiarity bias could risk the validity and reliability of phonemic transcription (Oller & Ramsdell, 2006). Therefore, it was deemed crucial to find an inter-transcriber who is not only trained in phonemic transcription, but who is also familiar with the Arabic and English phonemic systems. A Saudi PhD candidate in Phonetics and Phonology was judged to be a suitable candidate. The inter-transcriber is not only an experienced phonetician but is also a near native English speaker and a native Arabic speaker who is familiar with the Saudi Arabic dialect. Following a practice session involving 10 decoding outputs, the researcher responded to all questions raised by the inter-transcriber. Having ensured the inter-transcriber's understanding of the transcription procedure, she was then asked to transcribe 11 randomly selected cases. The same analytical method was then applied on the transcription data (see Appendix K).

Inter-transcriber reliability figures were 96.8% and 93.1% for the real word and pseudoword decoding outputs, respectively. However, again there was concern that the intra-transcriber reliability figures might be overestimated by the inclusion of correct items, therefore, the analysis was re-run after excluding these items, and the average inter-transcriber agreements were 86.3% and 86.9% for the real word and pseudoword decoding measures, respectively.

To conclude this section, both a strict and a more lenient analysis of intra- and inter-transcriber reliability suggest that the transcription of the decoding data is highly reliable. For a phoneme-level analysis, Shriberg and Lof (1991) suggest that, using broad phonemic transcription, inter-transcriber agreement figures of 85% and above are considered acceptable.

3.4.2.2.Interview A: The decoding experiences of Saudi foundation year ESL students

As mentioned in the above section, audio-recording the decoding sessions, and subsequent analysis of the decoding outputs, would allow an understanding of the decoding difficulties experienced by Saudi foundation year ESL students. Nonetheless, there are a few limitations to this error analysis method which only became apparent during the decoding sessions of the main study.

Firstly, given TOWRE's administration guidelines, students were allowed to skip (avoid) items which they found difficult to read. This avoidance was particularly observed among poor

decoders (Resorting to avoidance when encountering difficult structures is a widely-recognised limitation of Error Analysis (Ellis, 2008)). Secondly, as TOWRE is a speeded test, a speed-accuracy trade-off is not unexpected. That is, students might have produced inaccurate realisations which might otherwise be accurate (or inaccurate but in a different way) where the test was self-paced. Thirdly, as testing stopped when the 45 second period elapsed, not all students were able to read the same number of items, as more skilled decoders tended to read slightly more words than their less skilled counterparts.

These limitations could somewhat influence our understanding and representation of the difficulties that these students encounter. Therefore, the researcher decided to complement the error analysis data with short one-to-one interviews aiming at understanding students' experiences with English decoding. Interviews are useful for (a) validating other data-collecting methods (error-analysis data in the present study); and (b) obtaining information on humans' experiences (Cohen, Manion, & Morrison, 2011), thus rendering them a suitable tool for this part of the project.

As the interview aimed to complement the error-analysis data, the interview schedule consisted of one main question: "Tell me about your experiences of English decoding?" (a simpler Arabic translation was used for the specialist term 'decoding', one that corresponds to "pronouncing written words"), with a series of follow-up questions. Follow-up questions included "How do you find English decoding?", "What things do you find difficult about decoding English?", "How do you find the English script?". Then, following a free recall of participants' decoding experiences, the participants were shown a reading passage and were asked to inform the researcher if there were any difficult- /easy-to-read words. These were followed by discussions about why these words were found difficult/easy. This stimulated recall of their experiences guarded against any natural limitations on memory, a limitation of interview data.

A subsample of 12 students (A1=4, A2=4, B1=4) participated in this interview. The interviews took place in one of the university rooms and were all conducted in Arabic. Interviews were then translated into English by the researcher. A near native Arabic/English speaker was also consulted to assist with the translation of some colloquial phrases.

3.4.2.3.Think-aloud reports

The decoding tests would allow us to know whether or not, and the extent to which, Saudi foundation year ESL students encounter difficulties with decoding their second language. They would also allow us to analyse the decoding realisations which, together with the follow up interview above, would enable an understanding of the nature of their decoding difficulties.

However, as mentioned previously, tests offer limited insights into the processes that contributed to the obtained decoding realisations and scores (Woore, 2010). Understanding the decoding processes that readers engage in as they decode English could provide further explanations for the attested realisations, and eventually, scores, as Greenberg, Ehri, and Perin (2002, p. 241) noted,

> Researchers may want to include think-aloud procedures while participants are engaged in [reading aloud and spelling] tasks. Although many clues are derived from behavioral observations, much more can be gained by eavesdropping on the thought processes that occur before the behavioral response.

One way for obtaining information about participants' decoding processes is contained in the above quote and it is to have readers 'think aloud' while (or immediately after) decoding. The think-aloud methodology is based on cognitive information processing theories and the assumption that individuals can report what is happening (concurrent think-aloud), or what has recently happened (retrospective think-aloud) in their short-term memory (e.g., Ericsson & Simon, 1993; Olson, Duffy, & Mack, 2018). Following the analysis of these reports, researchers could then understand the underlying mental processes of interest.

The think-aloud methodology has been widely used in reading comprehension research in general (e.g., Cavalcanti, 1984; Davis & Bistodeau, 1993), and has been adopted for use, relatively recently, in decoding research (e.g., see Erler, 2003; Woore, 2010). However, despite the theoretical support for their use, and their relatively economic administration, two drawbacks of think-alouds are often noted in the methodological literature (e.g., Ericsson & Simon, 1993; Olson, Duffy, & Mack, 2018).

The first 'limitation' relates to the 'verbalisability' of the underlying processes. Not all mental operations are verbalisable. Some operations are performed automatically (unconsciously) and, as such, information about how a given task is being executed might not be part of the content of participants' short-term memory (e.g., Ericsson & Simon, 1993; Olson, Duffy, & Mack, 2018). Consequently, no information about the execution of the task could be reported during the think-aloud protocol. Empirically, this automatic mental operation could present itself in verbalisations along the lines of: "I know it", "I just said it", etc. (Woore, 2010).

The second limitation relates to the completeness and accuracy of the verbalised report. These could be hindered by several factors including memory limitations, concurrent administration of the think-aloud protocol (potentially causing "reactivity", i.e., causing participants to change their usual performance, and hence the cognitive processes involved),

or simply by individual differences in readers' ability to verbalise what is going on in their minds (e.g., Ericsson & Simon, 1993; Olson, Duffy, & Mack, 2018).

Given the present context of decoding, the first drawback was seen as an asset rather than a limitation. This is because "automatic processing" *is* a decoding process that readers are thought to engage in when reading words, particularly ones they are familiar with (see the popular dual-route models of word reading (Coltheart, Curtis, Atkins, & Haller, 1993) (section 2.1.2.2.6.2). This is why, in the present study, both familiar and unfamiliar words were used as stimuli in the think-aloud protocols. This is also why students were reminded that producing comments along the lines of: "I just said it, I know it" was acceptable. The presence of these familiar words, and the "I know it" comments they may trigger, would not only allow for the reporting of a natural decoding process, but would also provide a window into the validity of think-aloud use in decoding research (for example, if readers were more likely to produce such comments when reading familiar words than when reading unfamiliar words then that observation could support the use of think-alouds in decoding research).

However, a number of precautions needed to be made to minimize the impact of the second limitation, namely the completeness and accuracy of the reports. Firstly, retrospective, rather than concurrent, think-alouds were employed in the present study. This is because the rapid (milliseconds) duration of the decoding activity could make concurrent verbalisation difficult; logically, then, using concurrent think-alouds could change the usual sequence of thoughts, which would potentially contaminate the report data, as Ericsson and Simon (1993, preface) noted:

> For tasks of very short durations (less than a few seconds) verbal reports given retrospectively are even likely to be more complete than concurrent reports, as the former would present no difficulties with verbalisations of rapid sequences.

Furthermore, readers were asked to think aloud immediately following the decoding of each item (rather than following the decoding of a list of items). Though this would not completely guard against limitations on memory, it could greatly minimize their impact, as Olson, Duffy, and Mack (2018) note:

> If very short texts or text fragments were used, the memory problems are not as great, and useful TOL [i.e., thinking out loud] data could be collected after the fact.

In addition, as L2 language deficiencies could risk the accuracy of the reported data, all think aloud sessions were carried out in Arabic, the participants' L1. Finally, the data was also

examined for any consistency in participants' think-aloud reports (e.g., if the decoding processes were similar across the participants). The presence/absence of these consistencies could provide further evidence for the validity, or lack thereof, of the obtained verbalisations.

A subsample of 15 participants took part in the think-aloud task which took place in one of the university rooms. The report was obtained right after the participant decoded each of the 24 items (12 real word items and 12 pseudoword items) randomly selected from the L2 decoding proficiency test (every seventh word on the real word test and every fifth word on the pseudoword list[13]). The discussions were conducted on an individual basis. The table in Appendix E summarizes the administration procedure employed.

3.4.2.4. Interview B: The contribution of L2 decoding proficiency to L2 reading comprehension: readers' experiences

As noted previously, in past reading research, the conventional answer to "what is the contribution of x to y" has often pertained to the magnitude of the contribution of x, as estimated via quantitative statistical methods such as simple/multiple regression (or structural equation modelling). This is also the main methodological approach adopted in the present study to investigate the contribution of L2 decoding proficiency to L2 reading comprehension. This approach is undoubtedly valuable; however, it has some limitations which were noted in the literature review chapter (please see section 2.1.2.2.6.4), and which the researcher briefly reiterates here: Firstly, given the somewhat heavy data collection load in this line of enquiry, researchers are often restricted to a small set of independent variables (predictors) and/or small sample sizes. Limited variables and/or sample sizes could potentially influence findings as the former could detect an inexistent effect (Type I error), and the latter could suppress an existent effect (Type II error). In other words, in the case of the former, one cannot be certain that the contribution of x is genuine and not that of another excluded (third) variable. In the case of the latter, the contribution of x may be statistically insignificant simply due to the low statistical power resulting from the small sample size. Thus, exclusive reliance on null hypothesis statistical testing may at times be misleading. Secondly, the conventional answer to "what is the contribution of x to y" has often pertained to the *magnitude* of the contribution of x, rather than the *nature* of that contribution. That is, researchers interested in finding out the contribution of a variable would test the participants on this variable (and the dependent variable), determine the magnitude of the X contribution statistically, and refer to an existing

[13] except for one real word item <waves> and one pseudoword item <shrattec>. <waves> was read incorrectly by almost all participants and it was of interest to understand readers decoding processes which might have contributed to their incorrect realisation. <shrattec> was used to substitute <dreff> as participants tended to geminate double consonants and that gemination might not be clear if <dreff> was selected.

theoretical stance to explain the nature of that contribution. Thirdly, though not of primary concern, quantitative approaches leave unanswered the question of what might account for any unexplained variance in the regression model tested.

Therefore, a complementary qualitative component to the research design was applied to address these limitations. For this purpose, semi-structured interviews comprising task-based discussions were used. This is because: (a) interviews could validate other methods of investigations (Cohen et al, 2011). As such, they would allow us to see whether word reading difficulties – and also L2 vocabulary knowledge, L2 grammar knowledge, and/or the cognitive/comprehension skills masked in the L1 reading measure – emerge as ones of the issues that interfere with L2 reading comprehension; (b) interviews could act as explanatory devices for relationships between variables (ibid) and, thus, could allow us to understand what mechanisms underlie the contributory role of L2 decoding proficiency to L2 reading comprehension; and (c) interviews could enable an understanding of the factors that might account for any unexplained variance in the regression model tested.

All in all, the interview would allow for deeper interpretation of the quantitative results obtained. Figure 7 below outlines how the above limitations were addressed using the interview in the current study, along with the interview questions. The interview took a semi-structured approach allowing for a degree of flexibility where needed, with additional follow-up questions being asked where necessary.

A subsample of 18 students participated in this interview. The interviews took place in one of the university rooms and were all conducted in Arabic. Interview responses were then translated into English by the researcher. A near native Arabic/English speaker was also consulted to assist with the translation of some colloquial phrases.

Warm up, e.g., "what do you want to major in when you finish the foundation year?" "Why did you choose this major?"		
A. Issues affecting reading comprehension performance	**Free recall of readers' experiences with reading comprehension**	This part addressed limitations (a) and (c) discussed above
	- "From your reading experience, what issues interfere with your understanding when reading English texts? For example, what things make your understanding difficult? What things make it easy?"	
	Stimulated recall using a reading comprehension passage	
	- "Here is a reading comprehension text, could you take a look at it and let me know how you find it? - What things interfere with your understanding? What things make your understanding easy? What things make it difficult?"	
B. The mechanisms underlying the contribution of decoding to reading comprehension	- Where decoding-related issues emerged in A above, follow up questions were asked, for example, "How does this interfere with your understanding?".	This part addressed limitation (b) discussed above

Figure 6. Interview 2 outline

3.5.Data analysis methods:

The analytical methods adopted in the first and second parts of the study are discussed in turn below:

3.5.1. Part 1: The decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students

The first part of the study aimed to understand the decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students. The analysis methods employed to address these issues are described in detail below.

3.5.1.1.The decoding abilities of Saudi foundation year ESL students

To understand the decoding abilities of these learners, the researcher reports the decoding success rate for learners at all three levels of proficiency. Descriptive statistics (primarily means and mean percentage of success) were considered in order to understand learners' L2 real word, pseudoword, and overall decoding proficiency at each stage of L2 development (overall L2 decoding proficiency is a composite measure of L2 real word and L2 pseudoword decoding proficiency). Percentage success was seen as a particularly informative metric in the present study given the use of a standard and comprehensive measure of decoding skill (as opposed to idiosyncratic measures of decoding, i.e., a sample of words/pseudowords prepared by the researcher). Additionally, the decoding test scores were also described by reporting their L1 grade equivalents using the data published in TOWRE's examiner manual (Torgesen, Rashotte, & Wagner, 2012). It was thought that this information would provide further detail on the decoding abilities of L2 learners throughout the course of L2 development. L1 grade equivalents are also useful for comparability of research findings particularly where other standardised decoding measures (for example, Woodcock, 1987) are used as they remain the same regardless of the decoding test used. One-way ANOVA was also conducted to see if there was any evidence of growth in decoding abilities throughout the course of L2 development, i.e., whether there are significant between-group differences in real word decoding, pseudoword decoding, and overall decoding proficiency of learners at the three levels of proficiency (though the data in the present study is not longitudinal, and as such inferences were made based on comparisons between the different proficiency groups).

3.5.1.2.The decoding realisations of Saudi foundation year ESL students

When analysing learners' decoding realisations, 3 sub-objectives were identified to be addressed. ***The first objective*** was to explore the decoding difficulties of these learners by means of finding out what error categories would emerge from an analysis of their decoding outputs. Further, the availability of decoding data from A1, A2, and B1 learners would allow us to see whether learners at different levels of proficiency exhibit similar/different decoding difficulties, which in turn could provide rich information about the course of decoding development in L2 learners. ***The second objective*** was to see whether there are any developmental qualitative (or mainly quantitative) differences in the decoding difficulties observed (i.e., whether the groups exhibit similar/different nature of decoding difficulties, and/or whether they exhibit the same nature of difficulties, but to varying degrees, i.e., error rate). Additionally, a ***third objective*** was to investigate further these difficulties by examining

how L2 learners realise any given problematic phoneme, with a view to see if any patterns emerge in the realisations observed. It is thought that the presence (or lack) of patterns could contribute to, and enrich, discussion of these decoding difficulties. As mentioned previously, of particular interest here was to see if any patterns emerge in learners' realisations of vowel graphemes, as these could provide some insights into the previously-reported "vowel blindness" hypothesis. In this study, it is argued that if this is the case (i.e., if learners ignore written vowels), then one would expect to find the realisations of vowel phonemes to be random and lacking in consistency. On the other hand, consistency in the nature of the vowel errors observed could challenge (at least) a strong version of the vowel blindness claim (i.e., that Arab learners are not attentive to English vowels/ build a lexicon based solely on consonants).

As mentioned previously, 60 cases (20 per proficiency level) were randomly selected for their decoding realisations to be transcribed using the International Phonetic Alphabet. The realisations were phonemically transcribed by the researcher and later digitised to be prepared for coding and analysis. As students did not all read the same number of words (within the 45 second period), two main data files were created for the coding and analysis stage. The first included the full phonological transcription data, hereafter referred to as the **FPT dataset**. The second included a subsample of the phonological transcription data, hereafter referred to as the **SPT dataset**, which included only those items that were read by *all* participants, thus excluding any items that were read only by a portion of the participants. The choice of which dataset to use depended on the objective of the analysis at any given point. As will be discussed below, the FPT dataset was used for all analyses except those addressing between-group differences where use of the SPT data was deemed more suitable. The analysis procedures performed to address the three objectives are described in turn below. A summary of the analytical procedures is provided in Figure 8.

To achieve the first objective (i.e., to identify the decoding difficulties of the whole group), a decision was made to analyse the full phonological transcription data (FPT data). This is because the objective here was to capture as many decoding difficulties as possible (and not to investigate any between-group differences). All the decoding difficulties were coded as they were encountered in the phonological transcriptions. The codes were then collated into main themes (for example, vowel-related difficulty, consonant-related difficulty), sub-themes (for example, short vowels, under the vowel theme) and further sub-themes (for example, /æ/, /e/, /ɪ/, /ʌ/, /ɒ/, under the short vowel sub-theme).

To achieve the second objective (i.e., to see if there are any between-group differences in the decoding difficulties observed above), a decision was made to run the necessary analyses on items that were read by all the three groups (i.e., the SPT data). This is because higher proficiency learners tended to read more words than low proficiency learners, which might undermine the validity of any between-group differences found. Two sets of analyses were performed to achieve the second objective: **the first analysis** aimed to see if there are any "qualitative" variations in the decoding difficulties observed (i.e., whether the groups exhibit different/similar decoding difficulties). To this end, the SPT data file was split into three files (one for each group). The data was then examined to see if different themes emerge for learners at different levels of proficiency. **The second analysis** aimed to see if there are any quantitative variations in the difficulties observed (i.e., whether different proficiency groups exhibit similar/different number of errors in each of the emerging categories of error). To achieve this, the data was subjected to a series of one-way ANOVA (or the non-parametric alternative Kruskall-Wallis) with proficiency level as the independent variable (with three levels: A1, A2, and B1), and the frequency of error as the dependent variable. The error categories tested comprised all main categories (for example, "vowels") as the small size of the SPT dataset did not allow carrying out ANOVA on the subcategories (Note that the number of decoding items was approximately halved in the SPT dataset).

To achieve the third objective (i.e., to examine *how* L2 learners realise any given problematic phoneme and also see if any patterns emerge in the realisations observed) the following steps were taken: First, the realisations of each problematic phoneme were coded as they were encountered in the phonological transcription data (with the code being the exact phoneme produced). For example, for the grapheme <v>, as in <vasp>, the incorrect realisation of /fæsp/ was given the code /f/. Similarly, the incorrect realisation of /wæsp/ was given the code /w/, and so forth. Second, all the data relevant to the emerging codes were gathered. (Following the above example, all instances of /f/ where <v> should be decoded were gathered under the code /f/. Similarly, all the instances of /w/ where <v> should be produced where gathered under the code /w/). Third, coverage of the codes was presented in percentages of the total realisations observed for that grapheme to see if any realisations emerge as more frequently occurring than others. For example, if /f/ was observed on 90 occasions, and /w/ on 10 occasions (and no other realisations were observed), the percentage frequency would be 90% for the realisation /f/ and 10% for the realisation /w/. To ensure that the analysis covers as many realisations as possible, the FPT dataset was used for this analysis.

Finally, it should be noted that some challenges were encountered during the coding process. For example, the realisations of some pseudoword items could well fall under more than one category. By way of illustration, according to TOWRE's marking criteria (Torgesen, Rashotte, & Wagner, 2012), the pseudoword containing the grapheme <oo>, such as <sploosh>, could be correctly decoded as "splʊʃ" or "splu:ʃ". Thus, an incorrect decoding of this grapheme could fall under the short vowel category of /ʊ/ or the long vowel category of /u:/. To maintain consistency, a decision was made to code these errors in the long vowel category. A similar challenge was encountered when coding errors relating to the pseudoword <vasp>. This item for example could be correctly decoded as /va:sp/ (Standard British English) or /væsp/ (American English). Thus, incorrect readings of this item could correctly fall under the long vowel category of /a:/ and the short vowel category of /æ/. Again, in such cases, the key criterion was consistency of coding and in the present study such realisations were consistently placed under the /a:/ category. Finally, it is worth noting that some realisations were allowed to exist under more than one category. This was often the case when an error was coded under one category and also the "whole-word error" category (for example, a reading of <red>, /red/ as /ri:d/, <read> was placed under both the short vowel category and the whole-word-error category. Figure 8 below summarises the analytical procedures applied to analyse the decoding realisations.

<table>
<tr><th colspan="3">What is the nature of Saudi ESL learners (CEFR: A1, A2, and B1) decoding realisations?</th></tr>
<tr><td rowspan="3">Sub-Objectives</td><td colspan="2">Analytical methods</td></tr>
<tr><td colspan="2">Preliminary stage:
- 60 cases (20 per proficiency level) were randomly selected for their decoding realisations to be analysed.
- Phonological transcriptions were digitised following initial pen and paper transcription.
- Two main data files were created: FPT and SPT</td></tr>
<tr><td>Dataset</td><td>Summary of analytical procedures</td></tr>
<tr><td>1) To identify the decoding difficulties of learners at the lower end of the proficiency scale (i.e., the whole group)</td><td>FPT</td><td>• Error coding analysis
- decoding difficulties of the whole group (N= 60) coded as they were encountered
- codes collated into main themes, sub-themes and further sub-themes.</td></tr>
<tr><td>2) To examine whether there are any quantitative or qualitative differences between the three proficiency groups in the decoding difficulties observed.</td><td>SPT</td><td><u>Qualitative developmental differences:</u> (i.e., whether the three groups of proficiency exhibit different/similar decoding difficulties)
• Error coding analysis
- The SPT data file was split into three separate groups.
- data checked to see if the codes emerging in each group of proficiency were similar/different.
<u>Quantitative developmental differences</u> (i.e., whether different proficiency groups exhibit similar/different number of errors in each of the emerging categories of error):
• A series of ANOVA
- ANOVA (or the non-parametric alternative) was performed with proficiency level as the independent variable (with three levels: A1, A2, and B1, N=20 in each case), and the frequency of each error as the dependent variable.</td></tr>
<tr><td>3) To examine the nature of the realisations produced for each problematic phoneme.</td><td>FPT</td><td>• Error coding analysis
- The realisations of each problematic phoneme coded as they were encountered.
- All the data relevant to each of the emerging codes gathered.
- Coverage of the codes presented in percentages of the total realisations observed for that phoneme.</td></tr>
</table>

Figure 7. A summary of the analytical procedures applied to analyse learners' decoding realisations

3.5.1.3.The decoding experiences of Saudi foundation year ESL students

Interview responses were inductively coded as they were encountered and emerged from the data. Coverage of the codes across the samples was represented in percentages to guard against researcher bias (for example, bias apparent in overweighting narratives that conform to the pattern the researcher is interested in, and underweighting those that do not), thus enhancing the validity of the interview findings (Maxwell, 1992)). Response categories which emerged from one or two participants only (of the 12 participants), were merged and subsumed under an "other" code, with the underlying sub-categories retained and reported in the "other category" section.

3.5.1.4.The decoding processes of Saudi foundation year ESL students

A total of 360 think-aloud comments were coded as they were encountered. On occasions, a comment was given a single code (for example, small-unit processing), and on others, several codes (e.g., small unit processing *and* large-unit processing). Response categories which were observed five times or less were subsumed under an "other" code and are reported in the "other category" section. To ensure clarity of reporting, a coding system was followed when reporting results, and specifically, samples of learners' verbal reports. Graphemes were represented between angle brackets (e.g., <h>), phonemes between slashes (e.g., /h/) and letter names (e.g., aitch) by using capital letters (e.g., H).

3.5.2. Part 2: The contribution of L2 decoding proficiency to L2 reading comprehension

The second part of the study aimed to investigate the contribution of L2 decoding proficiency to L2 reading comprehension. When analysing the data for this part of the study, the following six issues were investigated in turn:

(1) The relationships between overall L2 decoding proficiency, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension ability;

(2) The contribution of overall L2 decoding proficiency to L2 reading comprehension before and after controlling for the contributions of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability;

(3) The contribution of overall L2 decoding proficiency as a function of the proficiency level of the L2 learner;

(4) The independent contributions of L2 real word decoding proficiency and L2 pseudoword decoding proficiency to L2 reading comprehension;

(5) The contribution of L2 decoding proficiency to L2 reading comprehension: further insights from Dominance Analysis and Relative Weight analysis

(6) The contribution of L2 decoding proficiency to L2 reading comprehension: qualitative insights from the experiences of second language readers.

In what follows, the analytical methods employed to address each issue, along with the rationale for their use, are presented.

3.5.2.1. The relationships between L2 reading comprehension performance, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 decoding proficiency, L2 real word decoding proficiency, and L2 pseudoword decoding proficiency

At the outset, it was deemed essential to examine the bivariate relationships between the study variables for two reasons. First, bivariate correlations would aid in the construction of the regression models adopted as they, among other functions, would allow us to (a) screen for the presence of collinearity between variables; and (b) create a parsimonious model by excluding variables whose correlations with the dependent variable were found to be insignificant. Second, bivariate correlations would aid in the interpretation of the regression model findings. For example, where a variable ceases to be a significant predictor of L2 reading comprehension in the presence of other variables in the model, examining the bivariate correlations between the insignificant predictor and its correlates would provide clues as to why that predictor lost its predictive power. Prior to running the correlational analysis, the data was checked to see if it met the assumptions of the Pearson Correlation test, and where not, Spearman correlations were used.

Upon conducting the correlational analyses, the next step was to examine the contribution of overall L2 decoding proficiency to L2 reading comprehension. In the present study, it was deemed important to study, and report, the contribution of L2 decoding proficiency to L2 reading comprehension before and after the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability were controlled for. This approach would not only inform our understanding of the contribution of L2 decoding proficiency, but also of how different analytical approaches may affect this contribution and any subsequent conclusions drawn. Despite its value, however, such approach to analysis, while adopted in previous prediction studies of L1 reading comprehension ability (e.g., Carr and Levy, 1990),

is not typically seen in L2 reading research. A detailed rationale for the two analytical techniques is outlined in the next two sections.

3.5.2.2. The contribution of L2 decoding proficiency *before and after* controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability

In the present study, it was deemed useful to initially examine the contribution of L2 decoding proficiency to L2 reading comprehension before controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability. This is because it has been argued that higher-level reading processes (i.e., semantic processing, syntactic processing, and the comprehension processes indexed in the L1 reading comprehension ability measure) are dependent on the execution of lower-level processes (such as decoding). That is, a reader cannot understand meanings, use their grammar knowledge, or benefit from their transferrable L1 reading comprehension ability without some sort of decoding (Nassaji & Geva, 1999). As Jiang et al, (2012, p. 333) comment, "in reading research, the decision to run [hierarchical] regression models… is theoretically driven. Reading performance can be conceived as consisting of component skills that are hierarchically ordered: from lower to higher levels". Thus, while controlling for other reading comprehension subskills would provide a rigorous statistical test for the predictive power of decoding (Carr & Levy, 1990), a strictly bottom-up view of reading (such as that adopted by Jiang et al, 2012) argues that, theoretically, it might not be meaningful to do so. This might be particularly the case when other reading comprehension subskills show strong relationships with decoding, as is usually the case in reading research. When this is the case, any shared variance between decoding and any of these variables would be credited to the variables entered first in the model.

Evidently though this view is problematic for many reasons (see below). However, for the present, it is sufficient to acknowledge its presence in the literature and, although the researcher does not support this view, it is worth noting that studying the contribution of decoding to reading comprehension *before* controlling for other reading subskills would enhance discussion of the findings, as it would facilitate comparability of research findings with other studies. This is because in the handful prediction studies that included a decoding measure, decoding was either: (a) the sole variable in the model (e.g., Koda, 1998), or (b) included with other variables but, presumably influenced by this bottom-up approach, researchers do not report whether L2 decoding proficiency continues to be a significant predictor of reading

comprehension in their final model (i.e., when the contribution of decoding can be examined in the presence of all other variables).[14]

For the above-mentioned reasons, initially, a hierarchical regression analysis was implemented to examine the contribution of L2 decoding proficiency to L2 reading comprehension *before* taking into account the effects of other reading comprehension subskills. Overall L2 decoding proficiency (a composite measure of L2 real word decoding proficiency and L2 pseudoword decoding proficiency), L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability were entered as predictor variables, and L2 reading comprehension as a dependent variable. The sequence of variable entry is presented in Figure 9 below. Following an initial run of the model, the data was screened to see if it meets the assumptions of multiple regression analysis (normality, homoscedasticity, linearity, independence of residuals, and absence of collinearity among the model predictors). Various graphical and statistical techniques where implemented for this purpose. Where a judgment was made using graphical methods, an expert statistician was also consulted to ensure validity and reliability of inference.

Step 1	Overall L2 decoding proficiency
Step 2	L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability

Figure 8. Order of variable entry in hierarchical regression

The analytical rationale noted above is not without problems. First, it somewhat reflects a "bottom-up" view of reading, which is no longer accepted by reading researchers (for example, several studies have showed that the linguistic context influences decoding performance (Stanovich, 1980, 1986, 2000). Thus, it might not be theoretically meaningful to assess the contribution of decoding proficiency independently of that of other predictors of reading comprehension skill. Second, and most importantly, given the high correlations between decoding proficiency and other reading subskills (Nassaji & Geva, 1999), conclusions regarding the contribution of decoding which are based exclusively on the R^2 of the first model

[14] To illustrate further, as a hierarchical regression analysis generates multiple models, the first (step 1) would present the amount of variance in L2 reading comprehension accounted for by L2 decoding proficiency alone, and the second (or third, etc.) will present the amount of variance in L2 reading comprehension accounted for by all independent variables including L2 decoding proficiency. However, potentially influenced by this bottom-up approach to reading, researchers would report model 1 relevant statistics (particularly R^2) regardless of whether decoding continued to be a significant predictor in their subsequent models (e.g., see Jeon, 2011 study). This is because they are often concerned with the size of R2 change that variables entered in step 2 are responsible for, regardless of the beta coefficients of the variables entered early in the model, and their significance values.

(i.e., where decoding is entered) might not be reliable. In hierarchical regressions, the use of first step/model R^2 as an index of a predictor's contribution has been criticized. This is because any shared explanatory variance would be credited to the predictor entered first in the model, thus potentially leading one to misunderstand (and likely overestimate) a predictor's true contribution (Tonidandel & LeBreton, 2011). In other words, one cannot be certain that it was decoding proficiency that accounted for this R^2 value (contribution), and not the other variables that it highly correlates with. As mentioned previously, to date, it remains unclear whether, and the extent to which, decoding proficiency would predict L2 reading comprehension when the effects of other key L2 reading comprehension variables are controlled for (Nassaji, 2014) and this is what the current analysis aims to address. As Carr and Levy (1990) note, such analysis, although it may be biased against decoding, would provide the most stringent test for its contribution to reading comprehension performance, and is, therefore, the approach that the researcher supports in the present study.

In light of the above, a second hierarchical regression analysis was performed to examine the data from a different perspective. The analysis was performed to see if prediction of L2 reading comprehension is enhanced by the inclusion of decoding proficiency after adjustment for variations in L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability.

In this analysis, the sequence of variable entry was as follows:

Step 1	L2 vocabulary knowledge L2 grammar knowledge L1 reading comprehension ability
Step 2	L2 decoding proficiency

Figure 9. Order of variable entry in hierarchical regression 2

3.5.2.3. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner

As this analysis aimed to discover whether the size of the decoding contribution, if any, varies across learners at different levels of linguistic proficiency, a decision had to be made first as to (a) how to operationalise the linguistic proficiency variable (i.e, the moderator); and following from that (b) how to enter this moderator variable in the model (i.e., as a categorical variable using a mean/median split, or as a continuous variable).

To operationalise linguistic proficiency, two approaches were considered suitable. The first was to operationalise linguistic proficiency using institutionally-determined CEFR proficiency levels (i.e., A1, A2, and B1). The second was to combine learners' scores on the L2 vocabulary knowledge and L2 grammar knowledge variables, with the resulting score being an index of learners' linguistic proficiency. The latter approach was deemed more suitable for two reasons: (1) the researcher had concerns about the possibility of a confound between the dependent variable (L2 reading comprehension) and the moderator variable (linguistic proficiency, when operationalised according to CEFR levels). This is because the language placement test (according to which learners were described as A1, A2, and B1) included a reading comprehension test and the researcher wanted to avoid a situation where the moderator variable was influenced by reading comprehension skill (Shiotsu, 2010); (2) this approach has been widely adopted in the L2 literature to operationalise L2 proficiency (Zhang, 2012), i.e., "Syntax and vocabulary have often been treated as synonymous to L2 proficiency" (Shiotsu, 2010 p.194). Therefore, students' scores on the L2 vocabulary knowledge and L2 grammar knowledge variables were added to create a linguistic proficiency variable, which was then used as a moderator in the regression analysis.

Once the proficiency variable had been operationalised, the second issue was to decide whether to keep this moderator variable continuous or to perform a mean/median split in order to transform it into a dichotomous moderator. A decision was made to retain the continuous nature of the variable (Hayes, 2018, personal communication). This is because artificial categorization of a continuum has been heavily criticized in the statistical literature. The mean/median split is problematic as it treats cases that are similar as very distinct (i.e., cases close to the mean/median point) (Field, 2015; Hayes, 2013). It has also been argued that it reduces statistical power of tests and might in some situations lead to a type I error (ibid).

A moderated multiple regression was conducted using the PROCESS macro for SPSS developed by Hayes (2013; 2018). The PROCESS tool has four advantages: (1) it automatically centres predictors; (2) it automatically computes the interaction term; (3) where an interaction effect is present, the tool performs a follow-up simple slope analysis which would allow us to understand the nature of the moderation effect; (4) to produce a visual representation of the interaction, the tool generates a set of estimates of the dependent variable for various combinations of the independent variable and the moderator variable using the regression model estimates and then generates a plot of the interaction (Field, 2015).

The regression model included L1 reading comprehension ability, L2 decoding proficiency, L2 linguistic proficiency (the moderator variable), and the interaction term L2 decoding proficiency*linguistic proficiency. In this analysis, it would be meaningless to include the individual predictors L2 vocabulary knowledge and L2 grammar knowledge, since the linguistic proficiency variable, which is a composite measure of both, is already entered in the model.

3.5.2.4. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the decoding measure

The analytical techniques adopted here aimed to address the following:

A. ***The contribution of L2 pseudoword decoding proficiency to L2 reading comprehension***
   1. Examine the contribution of L2 pseudoword decoding proficiency to L2 reading comprehension, both before and after the contributions of L2 linguistic proficiency, and L1 reading comprehension ability have been partialled out.
   2. Examine whether L2 pseudoword decoding proficiency predicts any variations in L2 reading comprehension after additionally controlling for L2 real word decoding proficiency.
   3. Examine whether the contribution of L2 pseudoword decoding proficiency would vary as a function of the linguistic proficiency of the learner, and if yes, how.

B. ***The contribution of L2 real word decoding proficiency to L2 reading comprehension***
   1. Examine the contribution of L2 real word decoding proficiency to L2 reading comprehension, before and after the contributions of L2 linguistic proficiency and L1 reading comprehension ability have been partialled out.
   2. Examine whether real word decoding proficiency predicts any variations in L2 reading comprehension after additionally controlling for L2 pseudoword decoding proficiency.
   3. Examine whether the contribution of L2 real word decoding proficiency would vary as a function of the linguistic proficiency of the learner, and if yes, how.

As the analytical methodology adopted in this section is similar to that adopted in sections 3.5.2.2, 3.5.2.3 and 3.5.2.4, it is not discussed further here.

3.5.2.5. The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance Analysis and Relative Weight Analysis

A key issue when performing a regression analysis concerns the "relative importance" of the predictors involved. In L2 reading comprehension research (and SLA research more generally), researchers traditionally relied on regression coefficients (beta weights) to determine the relative importance of their predictors. Beta weights were also valuable metrics in the present study (the analysis performed in section 3.5.2.2 would generate the beta weights of each predictor which would allow an examination of the relative importance of these predictors, i.e., predictors with larger beta weights are considered more important than predictors with smaller beta weights (Field, 2015)). However, language-related skills, including those predictive of reading comprehension ability, are often intrinsically correlated. In the presence of correlated variables, beta weights may be unreliable indicators of "relative importance" (Tonidandel and LeBreton, 2011), resulting in "at best limited information and, in some cases, erroneous interpretation" (Kraha, Turner, Nimon, Zientek, & Henson, 2012).

Therefore, the researcher was interested in examining the relative importance of L2 decoding proficiency to L2 reading comprehension using two further analytical techniques. Before discussing these techniques, the researcher synthesizes the key criticisms surrounding the use of beta weights as "sole" measures of importance, in the context of correlated predictors.

1. Over-estimating/under-estimating predictors' importance

Johnson (2000) cautioned that relying on beta weights – in the context of correlated predictors – may result in an inaccurate estimate of importance. Particularly, beta weights may overestimate the relative weight of the predictor most highly correlated with the criterion variable and underestimate of the weight of the other predictors in the model. This is illustrated in table 5 below adapted from Tonidandel and LeBreton (2011):

Table 5. An example illustrating some of the problems associated with Beta weights as a measure of relative importance in the context of correlated predictors (adapted from Tonidandel and LeBreton (2011).

Correlations						Relative importance metric		
	Y	X1	X2	X3	X4	Beta weight	General dominance weight	Relative weight
X1	.30	1				0.026	0.038	0.038
X2	.30	.60	1			0.026	0.038	0.038
X3	.25	.60	.60	1		0.001	0.019	0.019
X4	.25	.60	.60	.60	1	0.001	0.019	0.019

As can be seen in the table above, beta weights suggest that predictors XI and X2 are 26 times more important than X3 and X4. This is despite the minimal differences in the correlations between these predictors and the criterion variable Y (r=.30 for X1 and X2, and .25 for X3 and X4, i.e., a 5-point difference). Nonetheless, given the collinearity among the predictor variables (r=.60), beta weights appeared to have credited most of the shared variance among these predictors to XI and X2, just because of their slightly higher relationship with the criterion variable (Tonidandel and LeBreton, 2011). On the other hand, other measures of importance (General dominance and Relative weights found in the last two columns), which will be introduced in the following section, appear to have resulted in a more reasonable estimate of importance.

2. Interpretation problems

It has been argued that beta weights should not be used as indices for predictor importance as, in the presence of correlated variables, their interpretation becomes problematic and "awkward" (Azen and Budescu, 2003, p.130). Beta weights are interpreted as the expected change in the criterion variable that is associated with a one unit change in the predictor variable, *holding all other predictor variables constant*. However, when independent variables are correlated, this interpretation may not be meaningful, as an increase/decrease in one independent variable will intuitively lead to a change in the other variables it correlates with (Azen and Budescu, 2003; Luo and Azen, 2013). Nonetheless, despite being "the least informative", beta weights remain the typical approach for assessing relative importance (Braun & Oswald, 2011 p.332), possibly because these weights are automatically generated in popular statistical software packages (Tonidandel & LeBreton, 2015). Although no gold standard measure for relative importance exists when predictors are correlated, some metrics have been argued to be more *meaningful* than beta weights, such as dominance weights and relative weights (Braun & Oswald, 2011) which will be discussed later.

3. Context dependency

In addition, beta weights often change depending on the subset of independent variables included in the model. Thus, they are in a sense context-dependent (Lorenzo-Seva, Ferrando, & Chico, 2010). In fact, in some circumstances, these coefficients may even change in sign (negative/positive) as variables are added to/removed from the model (Kraha, Turner, Nimon, Zientek, & Henson, 2012). Thus, the accuracy of these regression coefficients is largely dependent upon having the "correct" model (context), which is rarely, if ever, known (Pedhazur, 1997).

4. Limited definition of importance

It is worth noting that no universally accepted definition of importance currently exists (Azen & Budescu, 2003). However, because beta weights only represent the unique contributions of the variable when determining its importance (and dismiss any variance that is shared among the predictors), some have argued that they cannot be reliably used as a measure of importance (Johnson, 2000). According to Hoffman (1960), for example, "relative importance" would refer to the contribution of each predictor to the total predicted variance when the predictor is considered by itself *and* in combination with other predictors (Johnson, 2000; Johnson & LeBreton, 2004). In other words, the relative importance of a predictor would incorporate both the unique effect of that predictor *as well as* the effect it shares with other predictors (ibid). It has thus been suggested that an informative measure of importance would be one that incorporates both the unique and shared effects of each predictor (Johnson, 2000).

5. Computational issues surrounding the use of beta weights and the practice of 'eyeballing' the size of beta weights when rank ordering predictors by importance

Finally, where variables' beta weights emerge as statistically significant, researchers typically evaluate their relative importance by 'eyeballing' the size of the beta weight associated with each variable. Predictors with larger beta weights are considered more important than those with smaller weights. However, it is important to note that the p-values generated with these coefficients only inform us of whether or not the coefficients are different *from zero*. They do not inform us of whether the relative importance of one predictor is significantly different from that of *another predictor* (LeBreton and Tonidandel, 2011). As such, and in light of this practice, it has been argued that they cannot be reliably used to rank order predictors' by importance (ibid). However, this practice of simply eyeballing the size of the beta weights is not surprising. The output typically generated by common statistical software often does not incorporate any information regarding the statistical significance of the difference between any two beta weights (ibid). Thus, researchers are left to make judgements about "relative importance" by relying on the size of the beta weights (ibid).

Because of the issues noted above, calls have been made to avoid *sole* reliance on beta weights, and to employ *additional/supplementary* techniques, to *support* the interpretation of regression results (Braun & Oswald, 2011; Kraha, Turner, Nimon, Zientek, & Henson, 2012). In particular, two statistics, General Dominance Weights and Relative Weights, might overcome some of the difficulties associated with Beta Weights (LeBreton & Tonidandel, 2008). Dominance Analysis (Azen and Budescu, 2003) and Relative Weight

Analysis (Johnson, 2000), were particularly developed to evaluate predictors' relative importance *in the context of correlated predictors*. However, despite emerging consensus for their superiority over traditional metrics of importance (i.e., beta weights) (Lorenzo-Seva, Ferrando, & Chico, 2010), the techniques are yet to be used in SLA research generally, and L2 reading comprehension research more particularly (but see Mellard, Anthony, & Woods, 2012; Tighe & Schatschneider, 2014 who used Dominance Analysis in L1 reading comprehension). These two techniques have thus been used in the present study to support the interpretation of the regression results, particularly, those in section 3.5.2.2. The techniques are discussed in detail below.

- Dominance Analysis

Azen and Budescu (2003) proposed conducting a dominance analysis to index predictors' relative importance. Dominance analysis rank orders the relative importance of predictors by examining the additional contribution (change in R^2) which each predictor makes to the variance explained *across all possible subset models (i.e., all possible combinations of predictors)*. The general dominance weight for each predictor is then obtained by averaging the individual additional contribution/change in R^2 values obtained across all possible subset models[15].

To illustrate, consider the hypothetical example where 4 predictors combined (x1, x2, x3, and x4) explain 58% of the variance in a given Y. Where one is interested in investigating the relative importance of X1, the additional contribution that X1 makes in all possible subset models would be calculated, that is:

1. The null model, i.e., a model where no predictors are added
2. A model including X_2
3. A model including X_3
4. A model including X_4
5. A model including X_2X_3
6. A model including X_2X_4
7. A model including X_3X_4
8. A model including $X_2X_3X_4$

Then, to estimate the general dominance weight of X_1, the additional contribution that X_1 makes (to all of the above models) is averaged. The process is then repeated to obtain the

[15] Azen and Budescu (2003) introduced a more detailed classification of dominance, describing dominance weights as being complete, conditional, or general. However, as the purpose of this section is to estimate overall importance, the focus is on general dominance weights, as they are suited for this purpose (LeBreton & Tonidandel, 2008).

general dominance weights of the remaining predictors (X_2, X_3, and X_4) (see table 6 below for detailed illustration).

Table 6. An example illustrating how General Dominance Weights are calculated (adapted from Azen and Budescu (2003))

General dominance weights provide a number of advantages over traditional measures of importance (i.e., beta weights). First, dominance weights are unique in that the relative importance of predictors is studied in the context of all possible models (Azen and Budescu, 2003). Thus, the weights generated are less context-dependent than beta weights. As a result, general dominance weights are more able to take into account the impact of any collinearity patterns within the data. Furthermore, because dominance analysis relies on an 'all subsets' approach, the general dominance weight represents the average contribution of a predictor to a criterion, both on its own and when taking all other predictors in the model into account (i.e., the weights incorporate both unique and shared effects) (Azen and Budescu, 2003; Braun and Olswald, 2011). This not only makes the weights fit the broad definition of relative importance presented earlier, but also make them more meaningful and easier to interpret (Johnson, 2000) (LeBreton & Tonidandel, 2008). Finally, general dominance weight has one desirable statistical property. When the values of the individual dominance weights are summed (i.e., in table 6 above, these are the values in the last row), the resulting value is the R^2 for the full model containing all of the predictors (which is .58 in the above example) (Braun & Oswald, 2011). Because they sum to the model R^2, General Dominance Weights can be interpreted as a measure of effect size (Azen and Budescu, 2003).

However, one issue associated with generating general dominance weights concerns the convenience of their computation. Performing "all subset regressions" can be computationally inconvenient when one has a large number of predictors. Furthermore, the dominance analysis approach allows an identification of a predictor's relative importance by a visual examination of the general dominance weights. Predictors with larger dominance weights are interpreted as more important than those with smaller weights. In other words, the approach does not rely on null hypothesis statistical testing to determine the importance of the weights, or the magnitude of the difference between any two weights (predictors).

- Relative Weight Analysis

Another way to compare the relative importance of predictors in multiple regression is to perform a Relative Weight Analysis (Johnson, 2000). As dominance analysis, Relative Weight analysis was developed to investigate the relative importance of predictors in the context of

correlated predictors. Particularly, Relative Importance Weights correct for the effects of collinearity among predictors prior to computing the contribution of each predictor to the model R^2.

Tonidandel and LeBreton (2011) and Lorenzo-Seva, Ferrando, and Chico (2010) summarized the computation method of relative weights in four steps, as follows. (For full details on the calculation method, please see Johnson (2000) and (LeBreton & Tonidandel, 2008))

1. The original predictor variables (*X*) are transformed to predictor variables (*Z*). The transformation method applied ensures a high degree of correspondence between the original predictors and the transformed ones. Nonetheless, unlike the original predictors, the new (transformed) predictors will not be inter-correlated.
2. The criterion variable (*Y*) is regressed onto the orthogonalized (transformed) predictors (*Z*). This will generate the standardized weights for each *Z* (transformed) variable.
3. Each *X* variable is regressed onto the transformed (*Z*) predictors which will generate a set of standardized regression coefficients.
4. The relative importance weights (RIWs) are computed by summing the squared standardized regression coefficients obtained in step 2 and 3 above. For example, the relative weight for X_1 will be: $\varepsilon_1 = \beta_1^2\lambda_{11}^2 + \beta_2^2\lambda_{12}^2 + \beta_3^2\lambda_{13}^2$.

LeBreton and Tonidandel (2008) cited three reasons why Relative weights are attractive to use: (1) as the transformed variables are uncorrelated, the variance in the dependent variable attributed to each predictor variable can be assigned unambiguously; (2) Like dominance weights, Relative Importance Weights range between 0 and 1 and their sum equals the model R^2. Thus, these weights can be interpreted as percentages of the explained variance associated with each predictor and thus can function as a measure of relative effect size, making them intuitively meaningful; (3) relative weights can be computed relatively easily regardless of the number of predictors involved in the model.

- Dominance Analysis or Relative Weight Analysis?

It is important to note that relative importance weights and Dominance weights almost always yield very similar results (Lebreton, Ployhart, & Ladd, 2004). Evidence for this comes from simulation studies which showed that the average difference between the two weights was often in the third decimal place (ibid). The fact that these two techniques generate similar

results despite their differential computation methods, has been considered an important strength (Tonidandel and LeBreton, 2011).

Given the similarity in the results obtained, Dominance Analysis and Relative Weight Analysis can be used somewhat interchangeably (ibid). However, there may be instances where one technique may be preferred over another. For example, where there is a large number of predictors, relative weight analysis may be more suitable (Tonidandel and LeBreton, 2011). This is because dominance analysis depends on an all subsets regression approach and, as such, the analysis would be more onerous where the number of predictors is large. To illustrate, a five-predictor study would require 31 separate regression models, whereas a 10 predictor study requires 1023 regressions (Johnson, 2000). Aside from computational convenience, in situations where one is interested in the statistical significance of the importance weights (or the statistical significance of the difference between these weights), a relative weight analysis may be preferred as these can be generated as part of the relative weight analysis output (Tonidandel and LeBreton, 2011).

Nonetheless, relative weight analysis is conducted on the full model comprising all predictor variables and there might be situations where one is interested in the detailed output of an all subsets analysis (LeBreton & Tonidandel, 2008). For example, one piece of information that dominance analysis provides is whether a predictor is more important than another in: (a) each subset model (complete dominance); (b) models of the same size, on average (conditional dominance); or (c) all subset models, on average (general dominance). Azen and Budescu (2003) illustrate how these types of dominance can be used to identify cases of suppressor variables. Thus, some valuable information may be obtained from an all subsets analysis which would not be available with a relative weight analysis. Where one is interested in this type of information, dominance analysis would be more suitable (LeBreton & Tonidandel, 2008).

In the present study, the reasonable number of predictors involved, allowed for a relatively convenient computation of the dominance weights. Nonetheless, it was of interest to examine the statistical significance of the importance weights (and the difference between them), a feature that current outputs of Relative Weights Analysis incorporate. In light of this, and because performing both analyses would allow an opportunity for analytical triangulation, a decision was made to subject the data to both analyses: Dominance Analysis and Relative Weight Analysis. Dominance analysis was performed using SPSS and an Excel program provided to the researcher by Dr. James LeBreton (Tonidandel & LeBreton, 2011). Relative

Weight Analysis was carried out using R on the RWA-Web platform (see Tonidandel & LeBreton, 2015).

3.5.2.6.The contribution of L2 decoding proficiency to L2 reading comprehension performance: qualitative insights from readers' experiences

Interview responses were inductively coded as they were encountered, and emerged from the data. Coverage of the codes across the participants was represented in percentages. On several occasions, a participant produced a response which included several codes. Thus, the overall total percentage may exceed 100%. Furthermore, response categories which emerged from one or two participants only (of the 18 participants) were merged and subsumed under an "other" code, with the underlying sub-categories retained and reported in the relevant "other category" section.

3.6.The piloting phase

The researcher had no concerns over the suitability of the institutional tests, given that her previous experience in using them was positive. However, it was thought useful to pilot them, along with the other data collection methods, to make any necessary adjustments, help the researcher gain confidence in administering the qualitative data collection methods in particular, and to identify any logistical issues.

At the outset, the researcher asked two ESL teachers (one with a PhD in Applied Linguistics and the other with a CELTA and over 18 years of English language teaching experience) to review the Vocabulary Levels Test and provide feedback about its suitability to low-level foundation year ESL students. As expected, the two teachers judged the 5000 word frequency band and 10000 word frequency band subtests to be excessively difficult. (In fact, the researcher and one of the two teachers agreed that students would be unlikely to know any of the vocabulary items on these subtests). While the 2000 level subtest was judged to be accessible, there were mixed views over the 3000 level subtest, with one teacher judging it as excessively difficult and the other as somewhat accessible. Furthermore, the researcher showed the grammar measure to 3 expert ESL teachers (two with a Master's degree in TESOL, and one with a PhD in applied linguistics) and they all identified the measure as a grammar one, despite being uninformed of the underlying construct the test is intended to measure. The researcher further asked one experienced teacher (with a Master's degree in TESOL) to identify the grammar domains in each test item (e.g., present simple, past, etc.) and the teacher was able to identify these domains successfully. Two experienced teachers (one with a PhD in applied linguistics and the other with a masters in TESOL) reviewed the TOEFL test and judged it to

be excessively difficult for foundation year students (in fact one of the teachers commented that she found it difficult herself).

In addition, the researcher administered the quantitative data collection methods to 31 foundation year students at the HEI whose proficiency levels ranged from A1 to B1 (A1= 11, A2=10, B1=10) (The decoding measures were piloted with nine participants only, three from each proficiency level) [16]. The researcher observed students' performance on the tests and asked them to provide any feedback they had. Several adjustments were then made to the tests, and/or their administration method following piloting. For example, while all the tests appeared to be of appropriate difficulty (e.g., all learners managed to complete the tests within the proposed time, with no floor or ceiling level performance), students found the 3000-level subtest of the VLT to be very difficult, which supports the judgement of one of the teachers who reviewed the test. A decision had to be made then as to whether to keep the subtest or remove it and keep the 2000 subtest only. Because few items on the 3000 subtest were known to the participants (e.g., the item "apartment= a place to live"), the researcher decided to keep the subtest, in addition to the 2000 subtest, but to emphasize to the main study participants that they might find the 3000 subtest difficult and that they should not engage in blind guessing. Furthermore, the researcher noticed that a number of participants matched the items on the VLT with their definitions by drawing lines instead of indicating the item number. Therefore, the researcher decided to explicitly inform the main study participants to answer the test by writing the item number. Additionally, two participants noted that it was not comfortable seeing the reading comprehension passage on one page and finding some of its questions on a different page. This was taken into consideration in the final version of the test by presenting the multiple-choice options of each test item horizontally rather than vertically so that each passage and its questions can fit in one page. Finally, the researcher asked the participants to indicate if they had previously encountered any of the test items on the real word measure. As discussed previously, most of the items (read within the 45 second period) were previously encountered by the participants.

In addition, the researcher piloted the think aloud protocol with three foundation year students (one A1 student and two B1 students). The researcher noted that participants tended to engage in conversations with her when she was seated next to them, but less so when she was seated out of their sight (at the back of the room). Therefore, the researcher decided to be seated out of the participants' sight throughout the whole think-aloud session. Another issue

[16] Three students did not attend the second session where grammar and L1 reading comprehension were administered.

was that students appeared unsure as to whether it is acceptable to state that they do not have any thoughts. Therefore, a decision was made to revise the oral instructions so as to assure the participants that they can simply say that they have no thoughts if they found that to be an accurate representation of what was going on in their minds.

Furthermore, pilot interviews into the issues that interfere with learners' reading comprehension performance were conducted with three students (two A2 students and one B1 student). The interviews highlighted areas of overlap in some of the original interview questions and issues with the wording of others, so both were revised/removed in the revised interview schedule. Further, as the interview included task-based discussions, students tended to point at any problematic words/structures and say "this" without actually articulating those words/structures. Therefore, a decision was made to keep a written note of them and/or articulate these problematic words/structures. Together, these revisions eliminated unnecessary redundancies and enhanced the credibility of the interview findings.

As for the decoding experiences interview, as mentioned previously, the researcher made the decision to collect this data during the main study phase. Due to limitations on time then, a decision was made to conduct the pilot interview with one participant only. However, the researcher planned to consider the first two main study interviews as pilot interviews if any difficulties arose. Because the pilot interview, and the first two main study interviews went smoothly, there was no need to conduct further pilot interviews, and no changes to the interview schedule were made.

Finally, some logistical issues were also noted. For example, the researcher initially planned to administer the four group testing measures (L2 reading comprehension, L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability) in one session. This was initially decided because ESL classes last for 3 to 4 hours, and as such, one classroom visit would allow for the administration of the four measures which would save time. However, the researcher asked the students to indicate the point at which they felt fatigued. Several students expressed, or showed signs of, feeling fatigued following the L2 vocabulary knowledge (the second measure administered) which prompted the researcher to stop testing. Therefore, to ensure that fatigue did not affect test performance, a decision was made to split the group testing session into two sessions, and to administer two measures per session. Furthermore, when the researcher administered the questionnaire at the beginning of session 1, she found that some questions triggered time-consuming discussions, and that the data of almost all the questions were already available to the researcher. Because the researcher had to

administer the L2 reading comprehension and L2 vocabulary knowledge measures in approximately one hour, a decision was made to free the session time for these measures, and to administer the background questionnaire at the beginning of the decoding testing sessions. In the next section, ethical considerations are discussed.

3.7.Ethical considerations

The present study has been conducted in accordance with the guidelines outlined in the code of conduct of The British Educational Research Association (2011). Having obtained approval from the University of Oxford's Research Ethics Committee (Appendix I), an official request to collect data from the data collection site was first made to the Saudi Embassy in London, which acts as the administrative link between sponsored students and Saudi higher education institutions. The embassy then forwarded the request to relevant officials in the higher education institution (the data collection site). Written consent was then obtained from the Vice-Dean of the university's English Language Institute, and the Vice Chancellor for scientific research (Appendix J). Having been allowed access to the site, consent was then obtained from participating students by inviting them to sign a consent sheet (Appendix K). As it was expected that the participants would have limited English proficiency, the researcher, a fluent Arabic speaker, translated the information and consent sheet to Arabic. The sheet described the aim of the research, why students were invited to participate, and their role in this research. It emphasized that there are no risks in participating, that participation is voluntarily, and that students can withdraw at any time without having to give a reason for doing so. Furthermore, the paper tests were kept in a locked room and the test scores and audio files (decoding tests, think-alouds, interviews) were securely stored in a password-protected computer. The data was kept fully confidential and anonymized using numerical codes.

As well as abiding by these official guidelines, two ethical considerations are worth noting:

First, unlike the individual decoding session which was scheduled during students' break times, the two group testing sessions were scheduled in the middle of ESL class times and so might have caused a level of disruption to students' learning. To somewhat mitigate this disruption, the researcher ensured that (a) all participating teachers were ahead of, or at least on track with, the institutional pacing guide (i.e., not behind in the curriculum); (b) the group testing sessions did not take place where the students had an English test on the same day, as students might want to spend the classroom time preparing for the test with their teachers. It is also worth remembering that, apart from the Arabic reading comprehension test, the tests

administered tapped into students' L2 reading comprehension, L2 grammar knowledge, and L2 vocabulary knowledge, and as such could be considered as part of their ESL class activity.

Secondly, as previously mentioned, English classes last for 3-4 hours, and it was often the case that the researcher would visit the class after the first teaching hour. Thus, at times, non-participating students were in the classroom during the group testing sessions, yet, they were not able to benefit from any learning the tests of the present study might have allowed. To minimize the risk of any harm to these students, the researcher (or the classroom teachers) suggested that they spend the time doing ESL activities found in their English workbooks. Some students asked if they could have a copy of the English tests without taking part in the project and the researcher provided them with these where extra copies were available.

Finally, a number of measures were taken to benefit the participants as much as possible:

First, during the individual sessions, the researcher responded to any questions the students put forward once the session was completed: for example, "How can I read word x?", "What is the meaning of word x?", and even more general questions, such as "How can I improve my English?".

Secondly, as a gesture of appreciation, a "thank you" note and a cupcake was given to all students at the end of the individual decoding session (a thank you card and a box of chocolates was also given to the teachers who welcomed the researcher to their classrooms).

Thirdly, all students participating in the quantitative phase of the project entered a draw to win 500 Saudi Riyals, approximately the equivalent of £102. Similarly, all students participating in the qualitative phase of the project entered a draw to win the same amount. Although the researcher initially considered reimbursing the students for their participating in the project, this decision was later revisited for two reasons: (1) the minimum hourly wage might not be attractive enough for some students, and given limitations on funds available, increasing that fee was not possible; (2) based on the researcher's familiarity with Saudi culture, and conversations with fellow teachers, there was concern that introducing money would be counterproductive. To illustrate, students may decide not to participate so as not to appear in need of any financial benefits, thus saving their social face. In fact, the researcher did not find it comfortable mentioning even the draw, and possibly neither did the students, judging by a few reactive comments along the lines of "We are not participating for money" and a few disapproving facial expressions.

Fourthly, many students commented that they were participating for religious causes (making comments along lines of: "I participated for God's sake", "I helped you so God sends

me someone to help me when I'm in need of help", etc.). Thus, it seemed that the very act of participating was in a sense emotionally rewarding to these students. Indeed, based on the researcher's familiarity with this context, these emotional rewards may even be of greater value than rewards in the form of course credits or money.

Finally, it is worth noting that, when presenting the project to the students, the researcher informed them that the nature of the tests that they will take (particularly the L2 reading comprehension and L2 grammar knowledge tests) is very similar to that of the institute's English exams (i.e., the exams that they will take for the English course). Students appeared to be eager about this and it might have helped familiarise them with institutional tests.

4. Results I: The decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students (CEFR: A1, A2, and B1)

This chapter presents the findings of the first two research questions:

RQ1: How well do Saudi foundation year ESL students (CEFR: A1, A2, B1) perform on measures of English real word and pseudoword decoding?

RQ2: What is the nature of their decoding realisations, experiences, and processes?

The chapter is divided into four parts. The first part (section 4.1) presents the findings into the *decoding abilities* of Saudi foundation year ESL students. The second part (section 4.2) presents the results relating to their L2 *decoding realisations*. The third part (section 4.3) presents the findings of the analysis of their L2 *decoding experiences*. Finally, the fourth part (section 4.1) presents the findings relating to their L2 *decoding processes*.

4.1. The decoding abilities of Saudi foundation year ESL students (CEFR: A1, A2, and B1).

4.1.1. Descriptive statistics and ANOVA

The mean pseudoword decoding scores of A1 (N= 71), A2 (N= 89), and B1 (N= 52) students were 19.89 (SD= 7.7), 25.42 (SD= 8.7), and 30.54 (SD= 8.4), respectively (i.e., mean success rate of 30%, 39%, and 46% respectively). The mean real word decoding scores were 40.23 (SD= 8.8), 51.43 (SD= 8.2), and 59.25 (SD= 6.7), respectively (i.e., mean success rate of 37%, 48%, and 55% respectively). Overall, the mean decoding proficiency of the groups were 60.11 (SD= 15.4), 76.84 (SD= 15.9), and 89.79 (SD= 13.3) (i.e., mean success rate of 35%, 44%, and 52%, respectively). The real word decoding proficiency of these A1, A2, and B1 students yield L1 grade equivalents of 1.8, 2.5, and 3, respectively, and the pseudoword decoding proficiency scores yield L1 grade equivalents of 2.2, 2.5 and 3.2 respectively.

Three observations can be made from these numerical results: (1) at a descriptive level, students encounter more difficulties with decoding pseudowords than real words; (2) students at all three stages of development encounter difficulties with second language decoding; and (3) decoding proficiency seems to increase throughout the course of L2 development.

To test the third observation further, i.e., to ascertain whether score differences between the three stages of development were statistically significant, the data was first checked for

normality, outliers, and homogeneity of variance. Given that the sample sizes were over 50 in each group (N= 71, 89, and 52 for A1, A2, and B1 learners, respectively), skewness and kurtosis z scores were used to check for normality in each group. As shown in Table 41 (Appendix L), the real word, pseudoword, and overall decoding proficiency scores were normally distributed for all three groups, as z values did not exceed ±3.29. To check for homogeneity of variance, the Levene's test was run. As shown in Table 42 (Apendix L), the groups have homogeneity of variance. Finally, to check for outliers, z scores were calculated for each group. Where a case had a z score that lay outside the ± 3.29 range, it was regarded as a potential outlier (Tabachnik and Fidell, 2014). No cases were identified as outliers.[17]

As the assumptions for the ANOVA were met, ANOVA was run with proficiency level as the independent variable and L2 real word decoding proficiency, L2 pseudoword decoding proficiency, and overall L2 decoding proficiency as the three dependent variables. Results showed that there was a statistically significant difference in the **L2 real word decoding proficiency** between the different groups, $F(2, 209) = 87.103$, $p<.001$. A Tukey post-hoc test revealed that there was a statistically significant difference between the A1 and A2 groups ($p<.001$), A1 and B1 Groups ($p<.001$), and A2 and B1 groups ($p<.001$). Additionally, there was a statistically significant difference in the **L2 pseudoword decoding** proficiency between the different groups, $F(2, 209) = 25.257$, $p <.001$. A Tukey post-hoc test revealed that there was a statistically significant difference between the A1 and A2 groups ($p<.001$), A1 and B1 Groups ($p <.001$), and A2 and B1 groups ($p<.001$). Finally, a statistically significant difference was also found in the **overall L2 decoding proficiency** between the different groups, $F(2, 209) = 59.488$, $p <.001$. A Tukey post-hoc test showed significant differences between all three groups ($p <.001$ for all pairs of comparisons).

[17] four of the 216 cases were identified as outliers but these four cases were removed anyway from this analysis, and therefore, no cases were further identified as outliers (of the 212 cases). A detailed discussion of these 4 outlier cases is provided in sections 5.1.1.3 and 5.1.1.4 in Results Chapter II.

4.2.The decoding realisations of Saudi foundation ESL year students

4.2.1. Categories of decoding errors

The analysis revealed four main categories of errors in the decoding realisations of the whole group (N=60). These are: (1) vowel-related errors; (2) consonant-related errors; (3) phoneme insertion, omission and mis-ordering errors; and (4) whole-word errors. These categories emerged in the whole sample analysis of both the FPT and SPT datasets (Figure 11), with largely similar proportional frequency, suggesting that the results observed remain robust across different samples of decoding items (2629 decoding items in the FPT dataset and 1510 items in the SPT dataset, though note that these figures include duplicates, i.e., items that were given two codes).

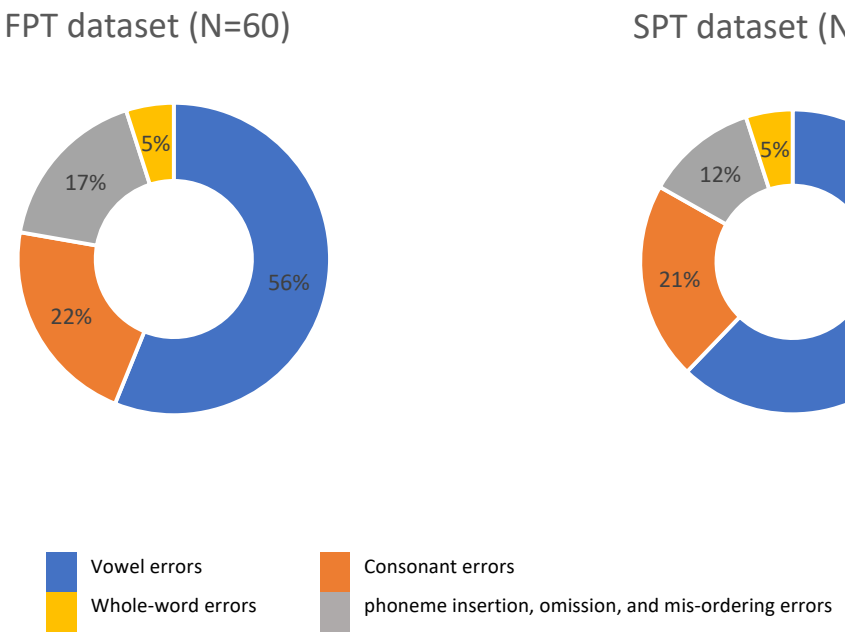


Figure 10. Categories of decoding errors in the whole-sample analysis along with their percentage frequency (of the total number of errors observed).

Furthermore, in the subsample analysis (Figure 12), the same four categories were observed in each group (A1, A2, and B1), suggesting that, at a macro level, no *qualitative* differences exist in the nature of decoding difficulties across these groups. It is worth noting too that the proportional frequencies of these categories were largely similar across the three groups. In other words, across the three groups, vowel errors accounted for the largest majority

of the total errors observed, followed by consonant errors, phoneme insertion, omission, and mis-ordering errors, and finally whole-word errors.

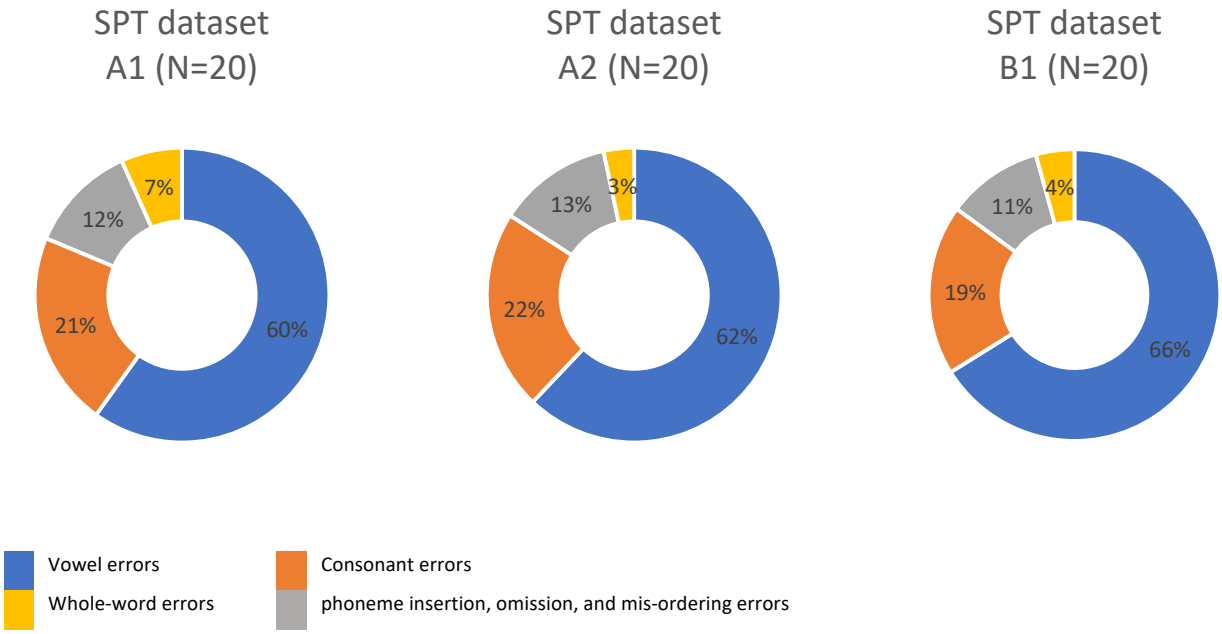


Figure 11. Categories of decoding errors in the subsample analysis along with their percentage frequency (of the total number of errors observed)

In what follows, results relating to each category of error are presented. Particularly, following a definition of what each category of error entails, two types of information are reported. The first deals with a *quantitative* between-group analysis of the error rates observed in the category in question. The second outlines the exact problematic phonemes found for each category together with the realisations produced for each. (This information is reported for the first two categories of errors, i.e., vowel related errors and consonant-related errors).

4.2.1.1.Vowel-related errors

An error was coded as vowel-related when the target vowel sound was substituted with an inaccurate one. As shown in Figure 11, these errors accounted for approximately two thirds of the errors observed in the whole sample analysis (of both the FPT and SPT datasets).

Furthermore, in a qualitative subsample analysis (Figure 12), vowel errors emerged in all three groups of proficiency (A1, A2, and B1), with a relatively similar proportional frequency. Now that vowel-related challenges were shown to be experienced by learners in all three groups (A1, A2, and B1), the next step was to see if these learners encounter these difficulties to similar/different degrees. To see if the vowel error rate varies across the three proficiency groups, the data was first checked for normality, outliers, and homogeneity of variance. Given the relatively small sample size (N=20), the Kolmogorov-Smirnov test was chosen to check for normality in each group. Results suggest that vowel error rates were normally distributed

for the A2, D(20) = .117, p = .200, and B1 groups, D(20) = .181, p = .085. However, the vowel error rate of the A1 group was not normally distributed, D(20) = .195, p = .045. To check for homogeneity of variance, the Levene's test was run. Results showed that these groups do not have homogeneity of variance, F(2,57)= 3.949, p = .025. Finally, to check for outliers, z scores were calculated for the vowel error rate for each group. Where a case had a z score that lay outside the ± 3.29 range, it was regarded as a potential outlier (Tabachnik and Fidell, 2014). No cases were identified as outliers.

As the normality and homogeneity of variance assumptions were violated, a non-parametric alternative of ANOVA, the Kruskal-Wallis test, was run. Results showed that there was a statistically significant difference in the vowel error rate between the different groups, $H(2)$ = 18.481, p = .000, with a mean rank vowel error rate of 42.50, 30.18, and 18.82 for the A1, A2, and B1 groups, respectively. Pairwise comparisons with adjusted alpha levels (for multiple comparisons) showed that there were no significant differences between the vowel error rate of the A1 and A2 groups (H= 12.325, p = 0.076) and A2 and B1 groups (H= 11.350, p = .118). However, the difference between the A1 and B1 groups was statistically significant (H=23.675, p = .000).

In what follows, the vowel phonemes which emerged as difficult, along with the realisations produced for each vowel, are presented. Because correctly decoded outputs were excluded from the transcription process, the realisations below cover the incorrect realisations produced for any given vowel. For organizational purposes, errors relating to short vowels are presented first, followed by long vowels, and finally diphthongs.

4.2.1.1.1. Short vowel phonemes

This section involves the realisations observed for any of the six English short vowel phonemes: **/æ/**, as in <cat>, /kæt/, **/ʊ/** as in <book>, /bʊk/, **/e/,** as in <red>, /red/, **/ɪ/,** as in <sip>, /sip/, **/ʌ/,** as in <cup>, /kʌp/, **/ɒ/,** as in <pot>, /pɒt/, and also, **/ə/** as in <mother>, /mʌð.ə/. Before presenting the realisations, however, the facility level for each of these short vowel phonemes is reported. Calculating facility level was not straightforward as no information was available on TOWRE's (both real word and pseudoword subtests') content validity (precisely, the frequency of occurrence of any given vowel phoneme). Therefore, to calculate facility level, a decision was made to select a random sample of one real word item and one pseudoword item for any given phoneme, calculate a percentage of success for each, and then average the obtained percentage of success. The sampling process was mostly run on the SPT dataset (to ensure the item was read by all the participants). However, where a vowel phoneme was not

present in the SPT dataset, the sampling process was re-run to comprise items from the FPT dataset. On the latter occasion, the number of participants who read that item was used as the denominator when calculating percentage of success. Table 7 below lists the facility level for all short vowel phonemes. The realisations produced for each short vowel phoneme are then presented.

Table 7. Facility level of short vowel phonemes

Phoneme	Pseudoword item	N	Phoneme percentage of success	Real word item	N	Phoneme percentage of success	Average percentage of success
/ʌ/	<w**u**m>	60	13.33	<j**u**mp>	60	96.67	55.00
/i/	<**i**p>	60	50.00	<h**i**m>	60	98.33	74.17
/ʊ/*	NA	NA	NA	<b**oo**k>	60	50.00	NA
/e/	<f**e**t>	60	26.67	<r**e**d>	60	66.67	46.67
/ɒ/	<p**o**th>	60	40.00	<l**o**st>	60	83.33	61.67
Notes: *No pseudoword item was found for this phoneme.							

- ***The realisations of the vowel phoneme /e/***

At the outset, it should be noted that in the dataset, /e/ was consistently represented using the grapheme <e> (for example, <red>) (n=193). As seen in Figure 13 below, a range of incorrect realisations were observed for this phoneme. Nonetheless, the proportional frequencies of these realisations were not equal, as some were clearly more frequently occurring than others, with /i/, /i:/, and /e:/ being the three most frequent realisations (accounting for 89% of the realisations) (see section 3.5.1.2 for a discussion of why examining the incorrect vowel realisations was of interest). It is also interesting to note that these most frequent realisations share a considerable degree of articulatory features with the target vowel (i.e., in terms of vowel height, backness, and lip rounding). That is, the target vowel is a high-mid front unrounded vowel, and the most frequent vowel realisations were mid to high front unrounded sounds, thus in very close proximity to the target vowel (see Figure 20 for an illustration). Furthermore, it is interesting that the second most frequent realisation /i:/ reflects the target letter name <e>.

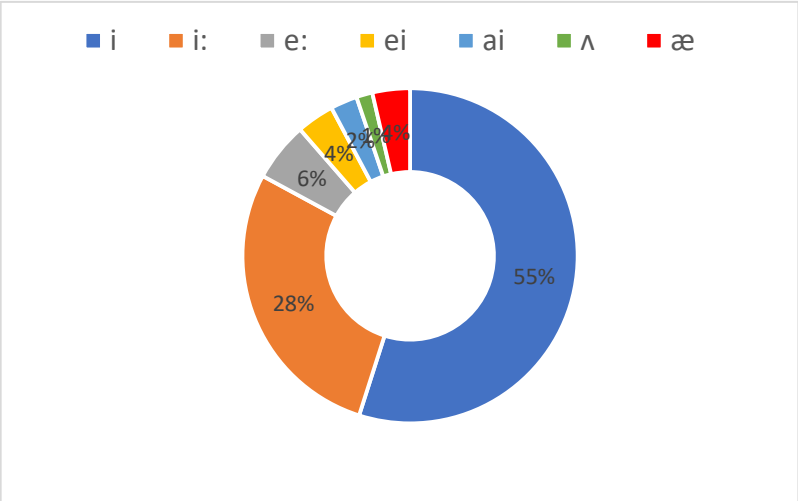


Figure 12. Realisations of the vowel phoneme /e/ and their percentage frequency

- ***The realisations of the vowel phoneme /ɪ/***

The phoneme /ɪ/ was almost always represented using the grapheme <i> (for example, <him>) (96.4%, n=244), except on 3.5% (n=9) of the occasions where it was represented in the grapheme <e> (for example, <depate>). As can be seen in figure 14 below, a range of realisations were observed for this vowel. Again, however, some realisations were more frequently occurring than others, with /ai/, /i:/, and /ei/ being the three most frequent realisations (accounting for 71% of the realisations). Again, the most frequent realisations, and particularly /i:/ and /ei/, appear to share a considerable degree of articulatory features with the target vowel (a high front unrounded vowel), as they are mid to high front unrounded vowel sounds. The sound /ai/, the most frequent realisation, does not share such a considerable degree of articulatory features with the target vowel, but, nonetheless, it is worth noting that it represents the letter name <i>.

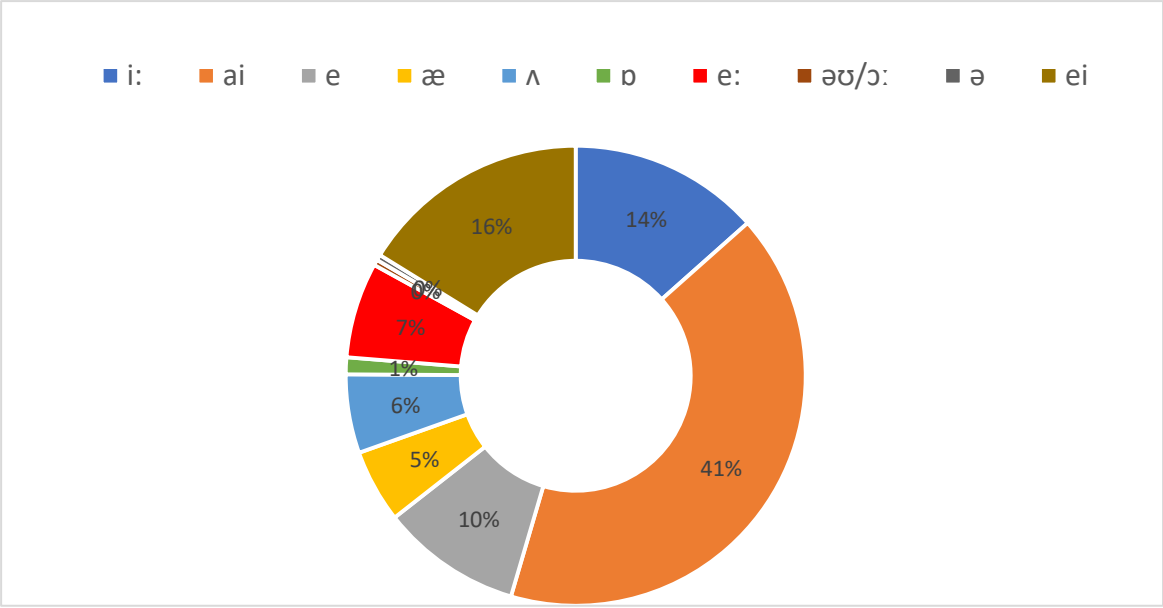


Figure 13. Realisations of the vowel phoneme /i/ and their percentage frequency.

As /ɪ/ was represented in two graphemes (<i> and <e>), it was of interest to examine the realisations observed for these two graphemic contexts separately. A detailed breakdown of the realisations obtained for each grapheme is presented in table 8 below.

Table 8. A breakdown of the realisations of /i/ across two graphemic contexts (<i> and <e>)

	realisation	Graphemic representation of /ɪ/	
		<i> (e.g.,<diss>)	<e> (e.g., <depate>)
1	**i:**	13.1%	22.2%
2	**ai**	42.6%	-
3	**e**	10.2%	-
4	**æ**	5.3%	-
5	**ʌ**	5.7%	-
6	**ɒ**	1.2%	-
7	**e:**	4.9%	55.6%
8	**əʊ/ɔ:**	0.4%	-
9	**ə**	0.4%	-
10	**ei**	16.0%	22.2%

Note: Some cells are unshaded to highlight that the realisations were present in one graphemic context only. Shaded cells highlight realisations which were present in more than one graphemic context.

A few points are worth noting from table 8 above: First, not all realisations were observed across the two graphemic contexts. In fact, the majority (70%) of the observed realisations were observed in a single context only, <i>. Second, where a realisation was present across the two contexts, its percentage frequency somewhat differed as a function of the graphemic

representation of /i/. The most prominent example here is /e:/ which was observed only 4.9% of the time in the <i> context, but 55.6% of the time in the <e> context. Third, the most frequent realisation in the overall analysis /ai/ (Figure 14) was never observed in the <e> context. Fourth, the three realisations observed for /i/ in the <e> context, /i:/, /e:/, and /ei/, also emerged as the three most frequent realisations for the vowel phoneme /e/ (which was represented in the grapheme <e>) (see the realisations of the phoneme /e/ in the preceding section).

➢ ***The realisations of the vowel phoneme /ʌ/***

The phoneme /ʌ/ was almost always represented using the grapheme <u> (for example, <jump>) (n=144) (except on one occasion where it was represented using the grapheme <o> in <money>). As can be seen in figure 15 below, a range of realisations were observed for this vowel. Nonetheless, some realisations were more frequently occurring than others, with /ɒ/, /ʊ/, and /ɔ:/ being the three most frequent realisations (accounting for 80% of the realisations). These frequent realisations do not appear to share many articulatory features with the target vowel (a low-mid back unrounded vowel) as they are all rounded vowels (and /ʊ/, and /ɔ:/ are also high vowels). They are however, like the target vowel, back vowels. Of the three realisations, /ɒ/ is slightly more similar to the target vowel, as it is also a low vowel.

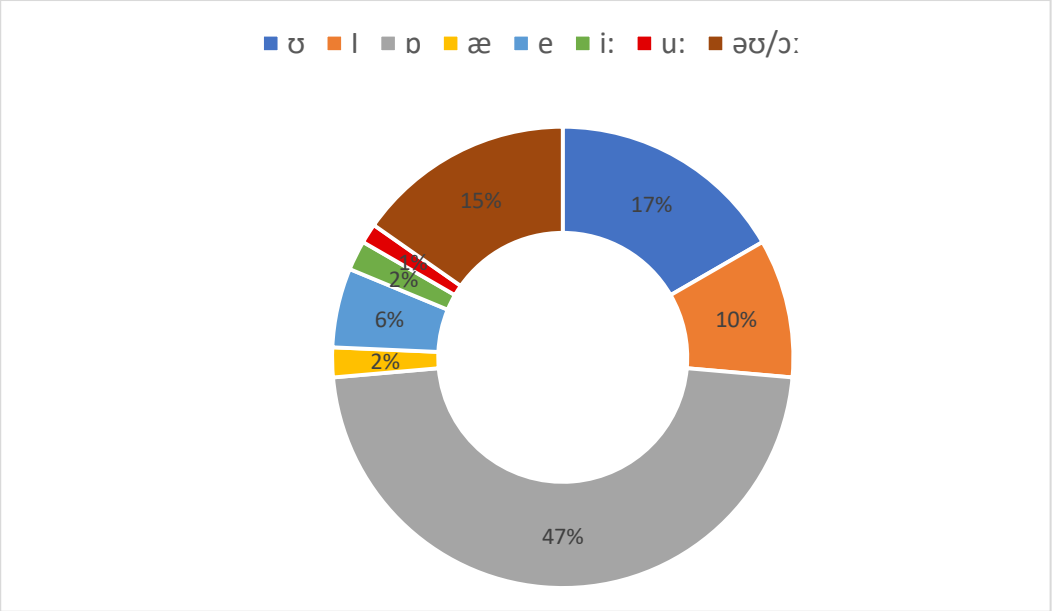


Figure 14. Realisations of the vowel phoneme /ʌ/ and their percentage frequency.

- ***The realisations of the vowel phoneme /ʊ/***

In the dataset, all instances of /ʊ/ were represented using the grapheme <oo> (for example, <book>) (n=64). As can be seen in figure 16 below, only three realisations were observed for

this vowel: /ɒ/, /u:/, and /ɔ:/. As with the realisations produced for the previous vowels, these also share certain articulatory features with the target vowel (a high back rounded vowel), as they are all mid to high back rounded vowels (except /ɒ/, in terms of height, as it is a low vowel).

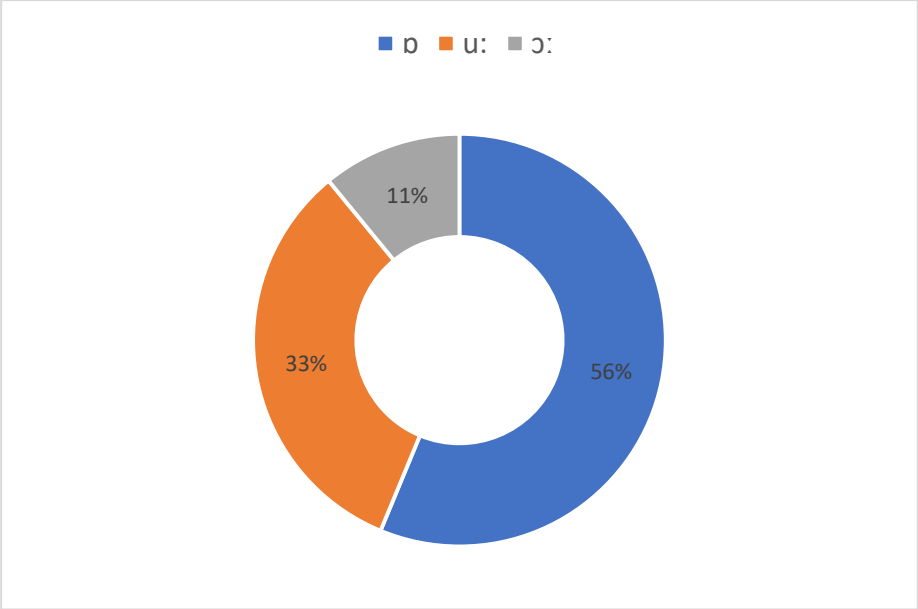


Figure 15. Realisations of the vowel phoneme /ʊ/ and their percentage frequency

- ***The realisations of the vowel phoneme /ɒ/***

The phoneme /ɒ/ was consistently represented using the grapheme <o> (for example, <lost>) (n=48). As can be seen in figure 17 below, a range of realisations were observed for this vowel. Again, however, (1) some realisations were more frequently occurring than others, with /əʊ/, /u:/, and /ʌ/ being the three most frequent realisations (accounting for 94% of the realisations); and (2) these frequently occuring realisations share certain articulatory features with the target vowel (a low back rounded vowel), as /əʊ/ and /u:/ are back rounded vowels. In terms of height, however, they are, unlike the target vowel, mid to high. The sound /ʌ/ is somewhat similar to the target vowel in the sense that it is a low-mid back vowel, although, unlike the target vowel, it is unrounded.

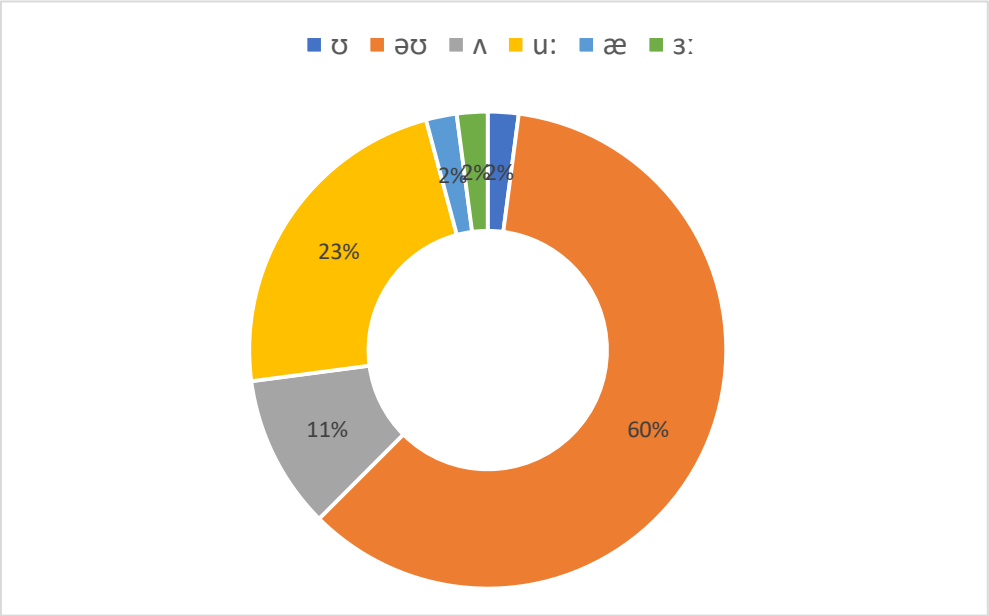


Figure 16. Realisations of the vowel phoneme /ɒ/ and their percentage frequency.

- ***The realisations of the vowel phoneme /æ/***

The sound /æ/ was consistently represented using the grapheme <a> (for example, <cat>) (n=57). As can be seen in figure 18 below, a range of realisations were observed for this vowel. Again, however, some realisations were more frequently occurring than others, with /ei/, /e/, and /ʌ/ being the three most frequent realisations (accounting for 76% of the realisations). These three realisations share some articulatory features with the target realisation (a low front unrounded vowel), as /ei/ and /e/ are both front unrounded vowels. However, in terms of height, they are mid to high. The sound /ʌ/, though a low-mid unrounded vowel, thus similar to the target vowel, is nonetheless different in the sense that it is a back vowel. Furthermore, it is worth noting that the most frequent incorrect realisation, /ei/, represents the letter name of the grapheme <a>.

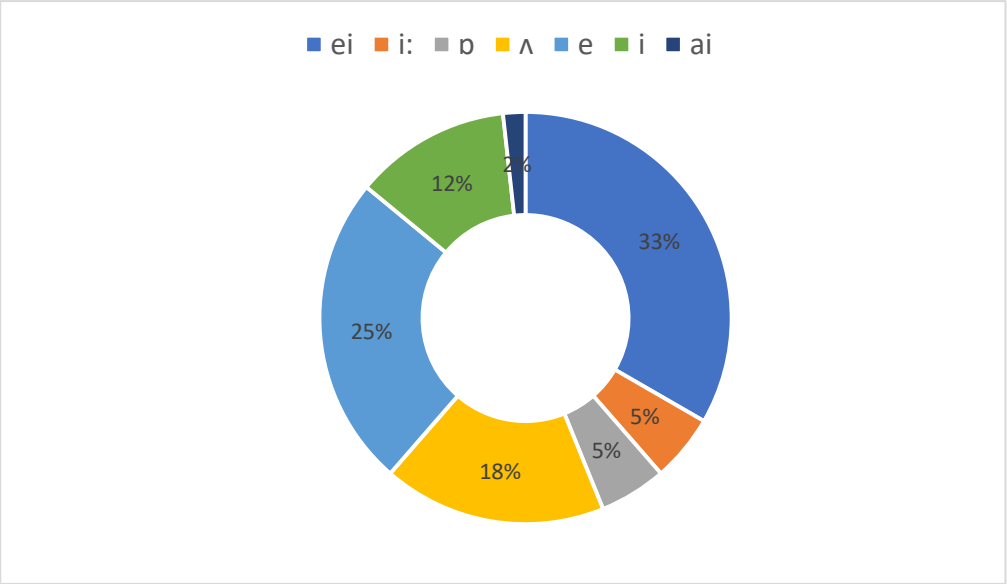


Figure 17. Realisations of the vowel phoneme /æ/ and their percentage frequency.

- ***The realisations of the vowel phoneme /ə/***

First of all, it should be noted that, acoustically, /ə/ does not differ (at least a great deal) from /ʌ/ presented earlier. However, from a transcription standpoint, it is standard practice to use /ʌ/ in stressed syllables and /ə/ in unstressed ones (for example, <mother>, /mʌð.ə/), hence their presentation in two separate sections here. As seen in figure 19 below, a range of realisations were observed for /ə/ and they share minimal articulatory features with the target vowel. Nonetheless, as the graphemic representation of /ə/ was more varied across the dataset, it was of interest to examine the realisations produced for each graphemic representation separately. Table 9 below outlines the graphemes used to represent this phoneme, the realisations obtained for each, and their percentage frequency.

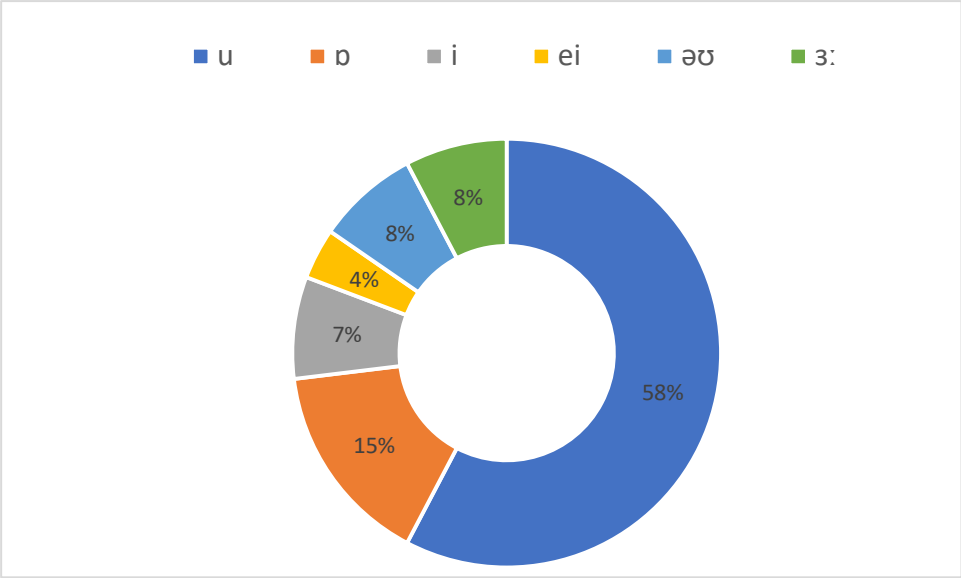


Figure 18. Realisations of the vowel phoneme /ə/ and their percentage frequency.

Table 9. A breakdown of the realisations of /ə/ across five graphemic contexts (<ou>, <o>, <er>, <e>and <->)

Realsiation	**Graphemic representation of /ə/**				
	<ou> (e.g,, <famous>)	<o> (e.g., <contain>)	<er> (e.g., <river>)	<e> (e.g., <the>)	<-> (e.g., <people>)
/ʊ/	100%	20%	0%	-	100%
/ɒ/	-	40%	33.3%	-	-
/əʊ/	-	40%	-	-	-
/e:/	-	-	-	100%	-
/i/	-	-	33.3%	-	-
/e/	-	-	16.6%	-	-
/ɜ:/	-	-	16.6%	-	-

Note: Some cells are unshaded to highlight that the realisations were present in one graphemic context only. Shaded cells highlight realisations which were present in more than one graphemic context.

Again, not all realisations were observed across all graphemic contexts (though, needless to say, the low frequency of the <e> and <-> contexts, (N=1), does not make this possible). Second, where a realisation was present in two contexts, its percentage frequency sometimes differed as a function of the graphemic representation of /ə/ (see shaded rows in table 9 above).

To conclude the short vowel section, all five English short vowels were observed in error analysis of learners' decoding outputs. Further, analysis of the nature of realisations produced for each vowel yielded a range of realisations. Nonetheless, there appeared to be little

evidence of randomness in the realisations produced given the considerable degree of consistency often observed. From a descriptive standpoint, this consistency was apparent in:

(1) the unequal distribution of the vowel realisations obtained for any given vowel phoneme, with some realisations occurring significantly more/less frequently than others;

(2) where a vowel phoneme had more than one graphemic representation, the observed realisations, and/or their proportional frequencies, tended to differ as a function of the graphemic context;

(3) the three most frequent realisations observed for each vowel phoneme – which accounted for an average of 84% of the total short vowel realisations – seemed to:

(a) differ as a function of the target vowel (table 10 below provides a summary of the three most frequent realisations obtained for each short vowel phoneme);

(b) share some articulatory features with the target vowel (i.e., in terms of vowel height, backness, and lip rounding) (Figure 20 below provides an example).

(4) On occasion, the most frequent realisations reflected the letter name of the grapheme used to represent the target vowel (e.g., the first most frequent realisations of the vowels /i/, and /æ/, which were mostly represented using the graphemes <i> and <a>, represent the letter names of these graphemes, i.e., /ai/ and /ei/, respectively. Further, the second most frequent realisation of the vowel /e/, which was mostly represented using the grapheme <e>, represents the letter name of this grapheme, i.e., /i:/).

Table 10. A summary of the short vowel phonemes and their three most frequently observed realisations

Target vowel	**Most frequent realisations**			
	1st most frequent	**2nd most frequent**	**3rd most frequent**	**Total % of the three most frequent realisations**
/e/	/i/	/i:/	/e:/	89%
/i/	/ai/	/i:/	/ei/	71%
/ʊ/	/ɒ/	/u:/	/ɔ:/	100%
/ɒ/	/əʊ/	/u:/	/ʌ/	94%
/æ/	/ei/	/e/	/ʌ/	76%
/ə/	/ʊ/	/ɒ/	/i/	81%
/ʌ/	/ɒ/	/ʊ/	/ɔ:/	80%
Average %				84%

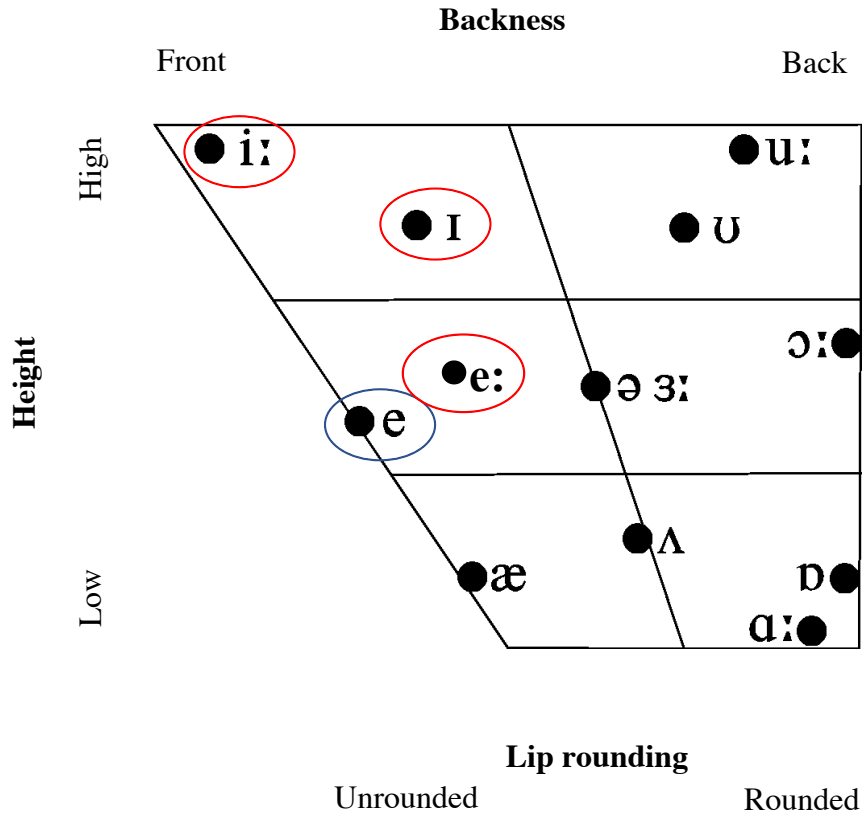


Figure 19. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target vowel /e/ (circled in blue). Note: */e:/ does not exist in RP.

4.2.1.1.2. Long vowel phonemes

This category involves any errors decoding the graphemes representing the five English long vowels: **/*a*ː/,** as in <far>, /f*a*ːr/, **/uː/,** as in <mood>, /muːd/, **/ɔː/,** as in <horse>, /hɔːrs/, **/ɜː/,** as in <birth> /bɜːrθ/, and **/i**ː/, as in <wheat>, /wiːt/. In what follows, the facility level of the long vowel phonemes are reported. Then, the long vowel phonemes which emerged as difficult are presented together with a description of the realisations produced for each.

Table 11. Facility level of long vowel phonemes

Phoneme	Pseudoword item	N	Phoneme percentage of success	Real word item	N	Phoneme percentage of success	Average percentage of success
a:	<v**a**sp>	60	80.00	<p**a**rt>	60	98.33	89.17
i:	<m**ee**st> ***	58	62.07	<h**e**>	60	98.33	80.20
u:*	<spl**oo**sh>***	31	50.00	<n**ew**>	60	86.66	68.33
ɔ:**	<k**o**>	60	85.00	<**a**lmost>***	54	61.11	73.06
ɜ:	<h**er**m>	60	86.67	<w**or**k>	60	13.33	50.00

Notes:
*For /u:/, only one real word item was found in the SPT dataset, <new>, and it was the item selected.
**for /ɔ:/, only one pseudoword item was found, <ko>, and it was automatically selected. Further, according to TOWRE's guidelines, three vowel realisaions were considered correct for this vowel (/ɔ:/, /əʊ/, and /ɒ/). Therefore, the accuracy figures for this vowel should be viewed with caution as the denominator (i.e., number of students who read the vowel correctly) does not only comprise correct readings of /ɔ:/, but also of /əʊ/, and /ɒ/.
*** Items were sampled from the FPT datasets, hence the smaller N reported.

The realisations for each vowel sound are reported below. The researcher stops to report here the exact articulatory similarities/differences with the target vowel and the reader is directed to the vowel chart for reference (figure 20).

- ***The realisations of the vowel phoneme /ɑ:/***

All instances of /ɑ:/ across the dataset were represented using the grapheme <a> (for example, <father>) (n=37). As can be seen in figure 21 below, a range of realisations were observed for this vowel, with /ʌ/, /i/, /e/ and /ɜ:/ emerging as the most frequent realisations (accounting for 86% of the realisations).

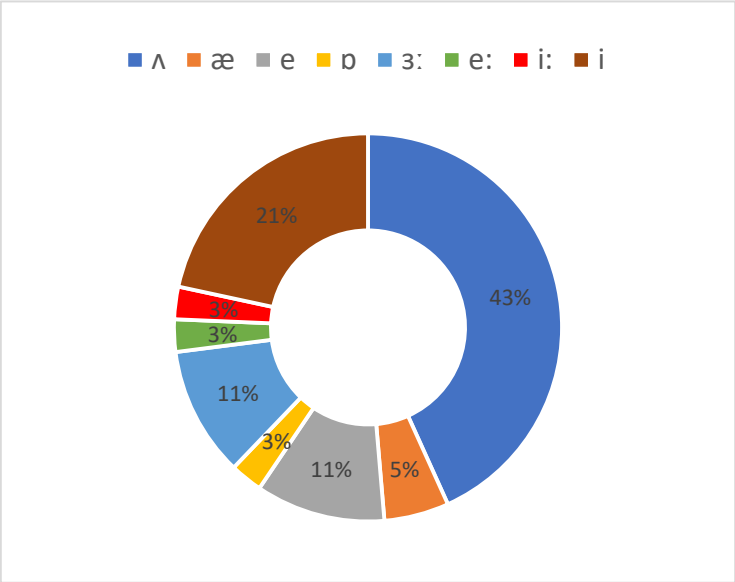


Figure 20. Realisations of the vowel phoneme /ɑː/ and their percentage frequency.

- ***The realisations of the vowel phoneme /iː/***

The phoneme /iː/ was almost always represented using the grapheme <ee> (for example, <meest>) (n=34) except for two instances (<people> (n=1), and <he> (n=1) where it was represented in the graphemes <eo> and <i>. As can be seen in figure 22 below, a range of realisations was observed for this vowel. Nonetheless, some realisations were more frequently occurring than others, with /i/, /ei/, and /e/ being the three most frequent realisations (accounting for 94% of the realisations).

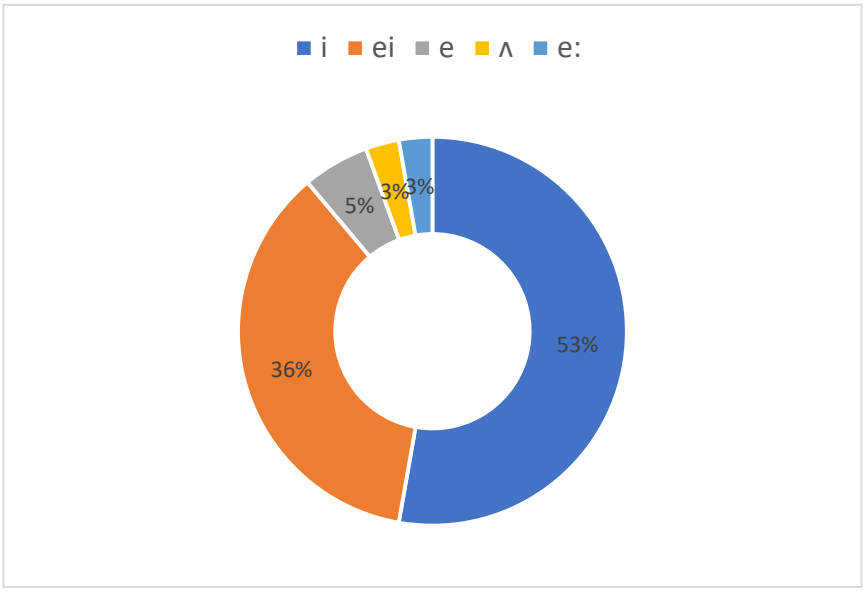


Figure 21. Realisations of the vowel phoneme /iː/ and their percentage frequency.

- ***The realisations of the vowel phoneme /u:/***

/u:/ was often represented using the grapheme <oo> (77%, n=38) (for example, <bloot>), but also in the graphemes <ew> as in <new> (16%, n=8), and <oe> as in <shoes> (6%, n=3). As can be seen in figure 23 below, five realisations were observed for this vowel. Nonetheless, some realisations were more frequently occurring than others, with /əʊ/, /ɒ/, and /aʊ/ being the three most frequent realisations (accounting for 96% of the realisations). A break down of the realisations obtained for each of the three graphemes, i.e., <oo>, <ew>, and <oe>, is presented in table 12 below.

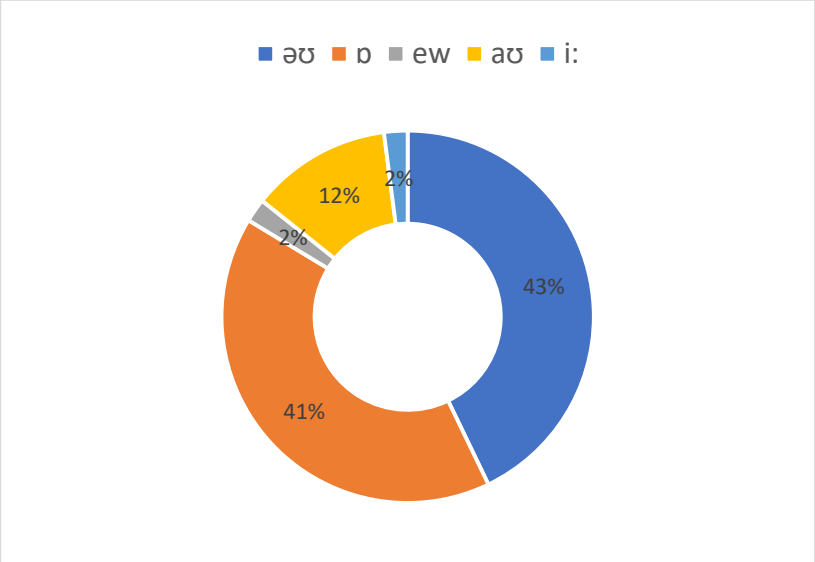


Figure 22. Realisations of the vowel phoneme /u:/ and their percentage frequency.

Table 12. A breakdown of the realisations of /u:/ across three graphemic contexts (<oo>, <ew>, and <oe>)

realisation	Graphemic context		
	<oo> (e.g., <sploosh>)	<ew> (e.g., <new>)	<oe> (e.g., <shoes>)
/əʊ/	44.7%	37.5%	33.3%
/ɒ/	50%	-	33.3%
/aʊ/	5.2%	50%	-
/ew/	-	12.5%	-
/iː/	-	-	33.3%
Note: Some cells are unshaded to highlight that the realisations were present in one graphemic context only. Shaded cells highlight realisations which were present in more than one graphemic context.			

Two points are worth noting from table 12 above: First, generally speaking, not all produced realisations of /u:/ were observed across the three graphemic contexts (except for /əʊ/. For example, the realisations /i:/ and /ew/ were never observed in the <oe> context.

Similarly, the realisation /ɒ/ and /i:/ were never observed in the <ew> context, and the realisations /aʊ/ and /ew/ were never observed in the <oe> context. Second, where a realisation was present in two or more contexts, its percentage frequency sometimes (though not always) differed as a function of the graphemic representation of /u:/. The most prominent example here is /aʊ/ which was observed only 5.2% of the time in <oo> context, but 50% of the time in the <ew> context.

- ***The realisations of the vowel phoneme /ɔ:/***

/ɔ:/ was represented using the graphemes <a> (as in <almost>) (70%, n=19) and <o> (for example, <ko>) (30%, n =9). As can be seen in figure 24 below, five realisations were observed for this vowel. Nonetheless, some realisations were relatively more frequently occurring than others, with /ə/, /u:/, and /æ/ emerging as the three most frequent realisations (accounting for 87% of the realisations). A breakdown of the realisations obtained for each of the two graphemes <a>, <o> is presented in table 13 below.

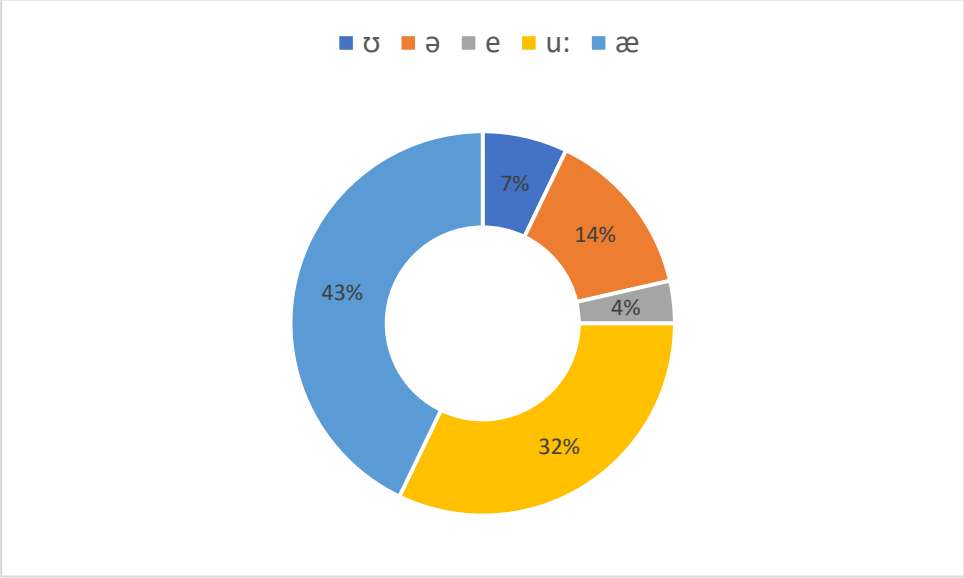


Figure 23. Realisations of the vowel phoneme /ɔ:/ and their percentage frequency.

Table 13. A breakdown of the realisations of /ɔ:/ across two graphemic contexts (<a> and <o>)

Realisation	Graphemic representation of /ɔ:/	
	<a> (e.g., almost)	<o> (e.g., ko)
ʊ	10.5%	-
ə	21.1%	-
e	5.3%	-
u:	-	100%
æ	63.2%	-

The table above shows that the observed realisations of /ɔ:/ clearly differed as a function of the graphemic context, as /u:/ was observed only in the <o> context, and all remaining four realisations (/ʊ/, /ə/, /e/, and /æ/) were observed only in the <a> context.

- ***The realisations of the vowel phoneme /ɜ:/***

In the dataset, instances of /ɜ:/ were represented using four graphemes <ur> (for example, <chur>) (25%, n=24), <ir> (for example, <clirt>) (11.5%, n=11), <er> (for example, <herm>) (9.4%, n=9), and <or> (for example, <work>) (54.2%, n=52). As can be seen in figure 25 below, a range of realisations were observed for this vowel, with /ɔ:/, /æ/, and /i:/ being the three most frequent relaisations (accounting for 82% of the realsiations observed).

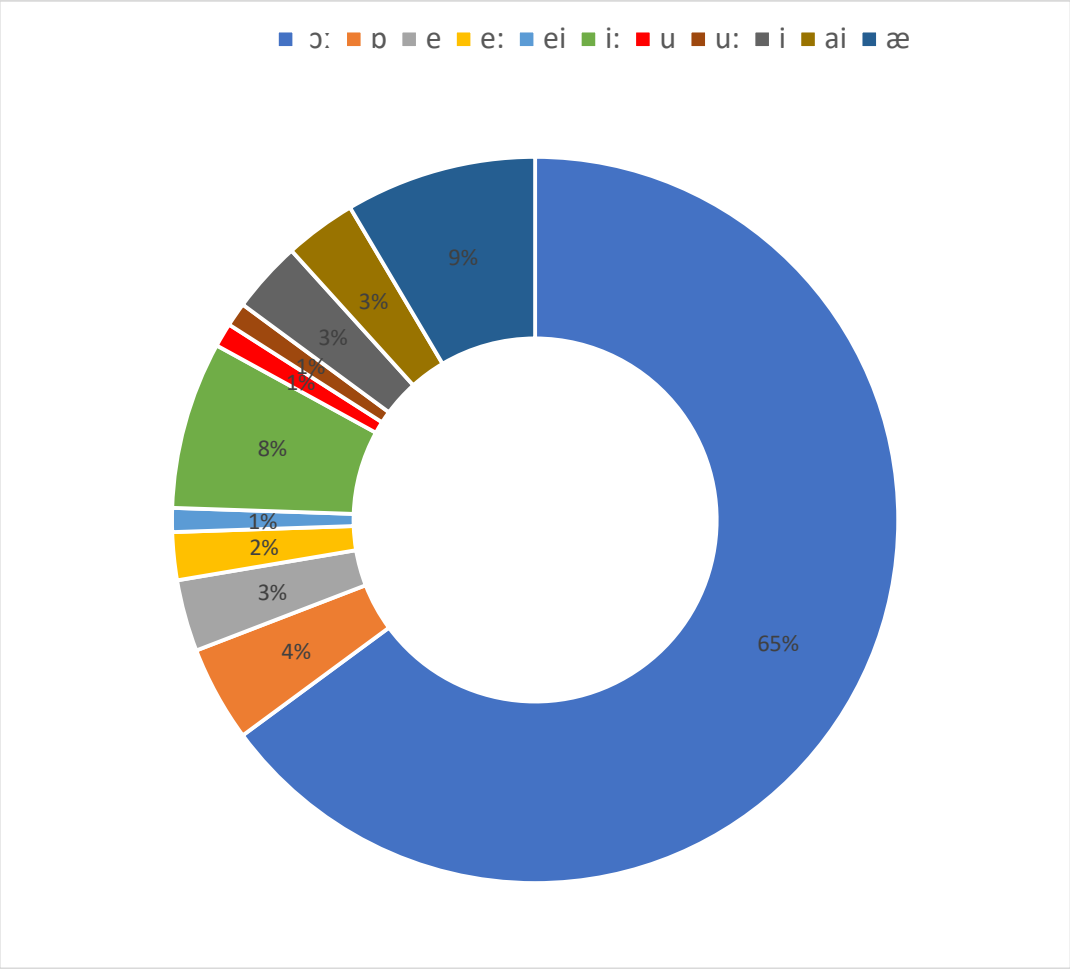


Figure 24. Realisations of the vowel phoneme /ɜ:/ and their percentage frequency.

Table 14. A breakdown of the realisations of /ɜː/ across four graphemic contexts (<ur>, <ir> ,<er>, and <or>)

	Realisation	Graphemic context			
		<ur>	<ir>	<er>	<or>
1	**ɔː**	50.0%	-	-	94.2
2	**ɒ**	-	-	11.1%	5.8
3	**e**	8.3%	-	-	-
4	**e:**	12.5%	-	-	-
5	**ei**	-	9.1%	-	-
6	**i:**	4.2%	18.2%	44.4%	-
7	**u**	4.2%	-	-	-
8	**ju**	4.2%	-	-	-
9	**u:**	4.2%	-	-	-
10	**i**	-	27.3%	-	-
11	**ai**	-	27.3%	-	-
12	**æ**	8.3%	18.2%	44.4%	-
13	**au**	4.2%	-	-	-

Note: Some cells are unshaded to highlight that the realisations were present in one graphemic context. Shaded cells highlight realisations which were present in more than one graphemic context.

Table 14 above suggests that the observed realisations of /ɜː/ clearly differed as a function of the graphemic context in that: (a) the majority of the realisations (10 out of 13) were observed in a single graphemic context only; and (b) where a realisation was observed across more than one graphemic context, the percentage frequency of that realisation appeared to vary as a function of the graphemic context (see shaded rows above).

To sum up the long vowel section, all five English long vowels were observed in the error analysis of learners' decoding outputs. Further, investigation of the realisations produced for each vowel showed a number of realisations. Again, however, there was evidence that the realisations were non-random given the following:

(1) the unequal distribution of the vowel realisations obtained for any given long vowel phoneme, with some realisations occurring significantly more/less frequently than others;

(2) where a long vowel phoneme had more than one graphemic representation, the observed realisations, and/or their proportional frequencies, tended to differ as a function of the graphemic context;

(3) the three most frequent realisations observed for each long vowel phoneme – which accounted for approximately 90% of the total vowel realisations – tended to:

(a) differ as a function of the target vowel (see table 15 below for a summary of the most frequent realisations obtained for each short vowel phoneme);

(b) share some articulatory features with the target vowel (i.e., in terms of vowel height, backness, and lip rounding, see Figure 26 below for an illustrative example).

Table 15. A summary of the long vowel phonemes and their three most frequently observed realisations

Target vowel	**Most frequent realisations**			
	1st most frequent	**2nd most frequent**	**3rd most frequent**	**Total % of the three most frequent realisations**
/ɑː/	/ʌ/	/i/	/e/ and /ɜː/	86%
/iː/	/i/	/ei/	/e/	94%
/uː/	/əʊ/	/ɒ/	/aʊ/	96%
/ɔː/	/əʊ/	/uː/	/æ/	87%
/ɜː/	/ɔː/	/æ/	/iː/	82%
Average %				89%

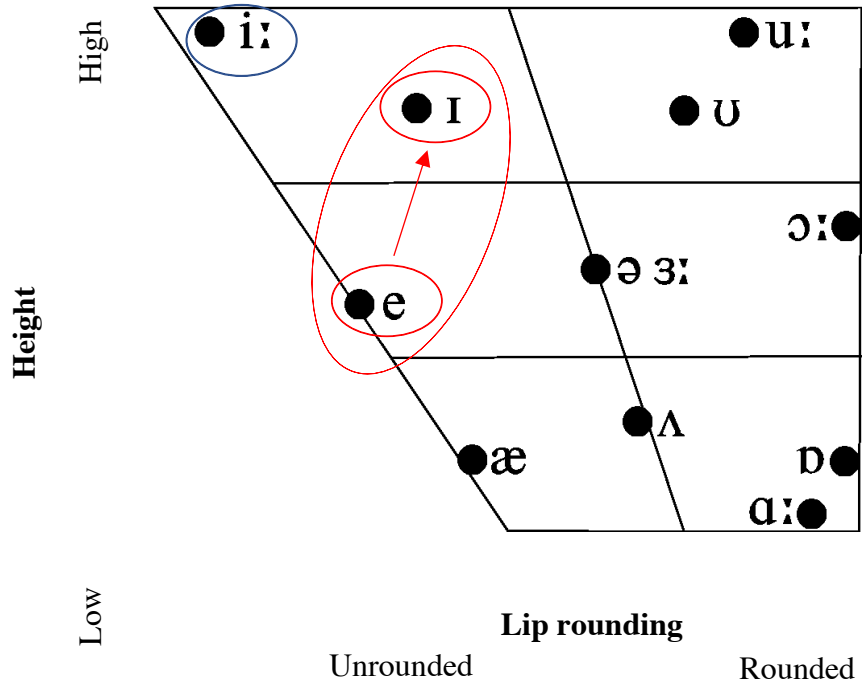


Figure 25. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target long vowel /iː/ (circled in blue). (note: the arrow indicates the diphthong /ei/).

4.2.1.1.3. Diphthongs

This category involves any errors decoding the graphemes representing the following diphthongs[18]: **/aɪ/**, as in <bike>, /bike/, **/eɪ/**, as in <late>, /leɪt/, **/aʊ/,** as in <down>, /daʊn/, and, and **/əʊ/** (or /oʊ/, the American equivalent of British /əʊ/), as in <road>, /rəʊd/. In what follows, the diphthongs which emerged as difficult are presented together with a description of the realisations produced for each.

Table 16. Facility level of diphthongs

Phoneme	Pseudoword item	N	Phoneme percentage of success	Real word item	N	Phoneme percentage of success	Average percentage of success
/aɪ/	<tive>	59	42.4	<kind>	60	71	56.7
/ei/	<bave>	60	48.3	<able>	60	41.7	45
/əʊ/	<strone>		85.71	<open>	60	38.33	62.02
/aʊ/	NA	NA	NA	<crowd>**	52	73.1	NA
Notes: *one case did not read this item. ** <crowd> was sampled from the FPT dataset, hence the relatively smaller number of cases who read this item.							

- ***The realisations of /aɪ/***

In the dataset, instances of /a**ɪ**/ were mostly represented using the graphemes <i> (for example, <kind>) (97.4%, n=113), but also in the grapheme <y> (for example, <sky>) (2.6%, n=3). As can be seen in figure 27 below, a range of realisations were observed for this vowel, with /i/, /ei/, and /i:/ being the three most frequent relaisations (accounting for 86% of the realsiations observed).

[18] These are the only diphthongs that were present in the items that the 60 cases decoded. For example, the diphthong /ɔɪ/ was present once in TOWRE in the item <boisterous>. Nonetheless, this item was presented toward the end of the test, and no participant managed to decode the item given the 45 second time limit.

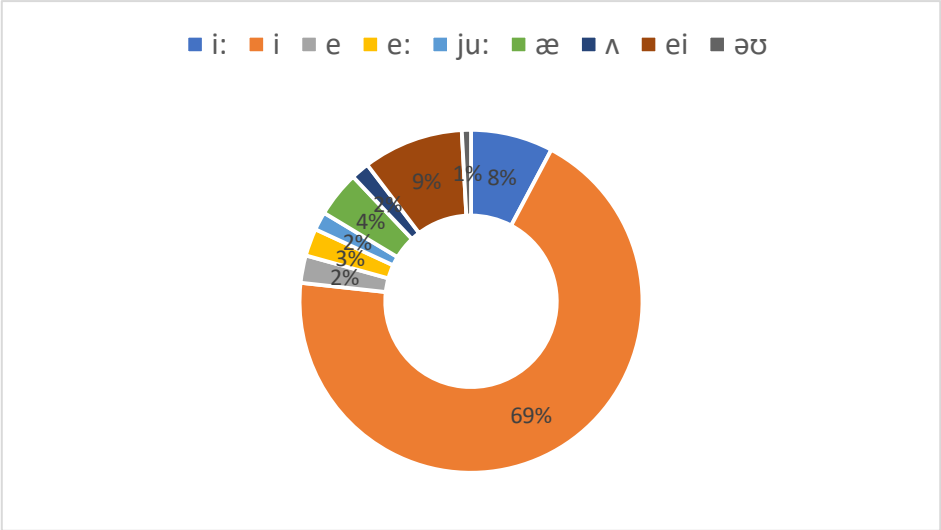


Figure 26. Realisations of the vowel phoneme /aɪ/ and their percentage frequency

- ***The realisations of /eɪ/***

The sound /eɪ/ was represented in the split diagraph <a-e> (for example, <pave>) (100%, n=229). As can be seen in figure 28 below, a range of realisations were observed for this vowel, with /æ/, /ə/, and /aɪ/ being the three most frequent relaisations (accounting for 68% of the realsiations observed). It is worth noting here that the realisation /ə/ was only observed in one word <**a**ble>.

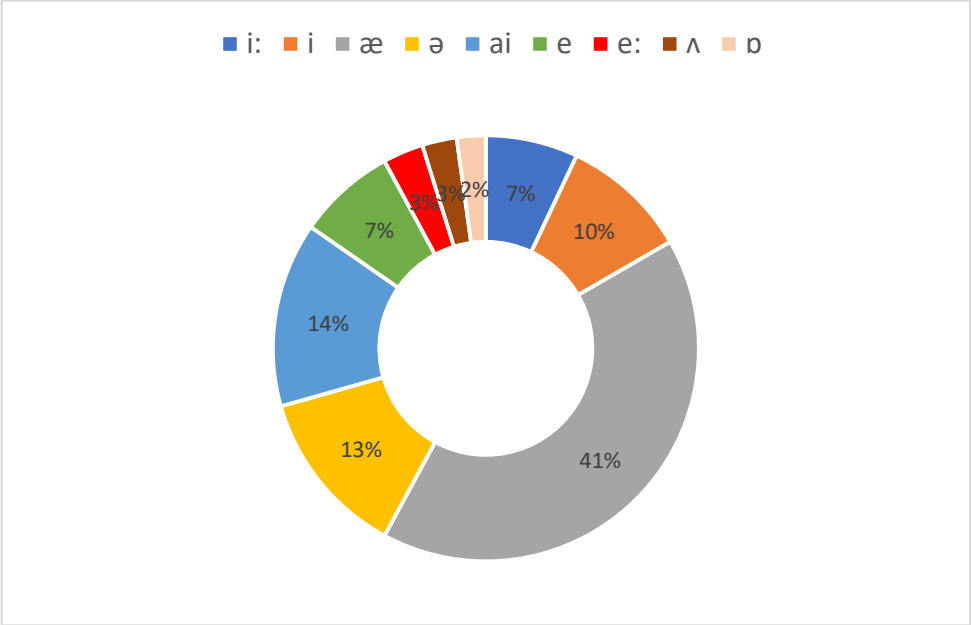


Figure 27. Realisations of the vowel phoneme /eɪ/ and their percentage frequency.

- ***The realisations of /əʊ/***

/**əʊ**/ was represented in the grapheme <o> (for example, <strone>) (100%, n=83). As can be seen in figure 29 below, a range of realisations were observed for this vowel, with /**ɒ**/, /**ɔ**:/, and /**ʊ**/ being the three most frequent realisations (accounting for 96% of the realisations observed).

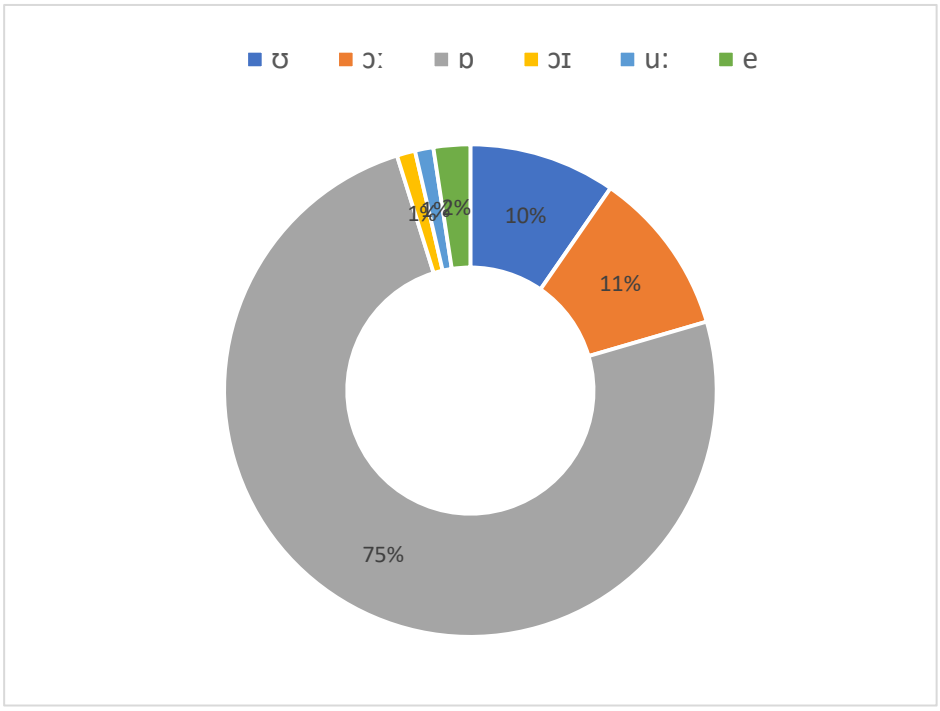
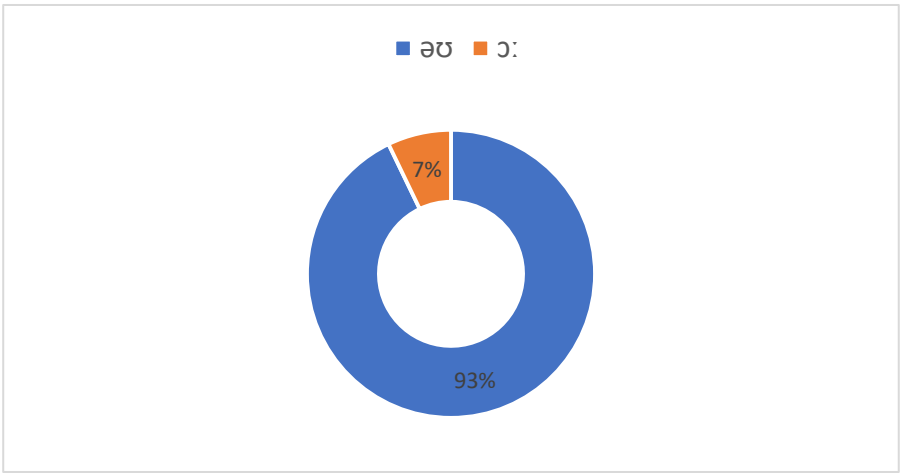


Figure 28. Realisations of the vowel phoneme /əʊ/ and their percentage frequency.

- ***The realisations of /aʊ/***

/**aʊ**/ was represented in the grapheme <ow> (for example, <crowd>) (100%, n=14). As can be seen in figure 30 below, only two realisations were observed for this vowel, /**əʊ**/ and /**ɔ**:/.[19]



[19] Owing to their similar articulatory features, the category /əʊ/ might have involved realsaitions of /o:/ and /oʊ/ (See appendix L)

Figure 29. Realisations of the vowel phoneme /aʊ/ and their percentage frequency.

To sum up the Diphthong vowel section, four diphthongs were observed in error analysis of learners' decoding outputs. Further, analysis of the nature of realisations produced for each diphthong yielded a range of realisations. Nonetheless, there appeared to be little evidence of randomness in the realisations produced. Similar to short and long vowels, this was apparent in:

(1) the unequal distribution of the vowel realisations obtained for each diphthong, with some realisations occurring significantly more/less frequently than others;
(2) the three most frequent realisations observed for each diphthong – which accounted for approximately 88% of the total vowel realisations – tended to:

(a) differ as a function of the target vowel (see table 17 below for a summary of the most frequent realisations obtained for each diphthong);

(b) share some articulatory features with the target vowel. (Figure 31 below provides an example).

Table 17. A summary of the target diphthongs and their three most frequently observed realisations

Target vowel	**Most frequent realisations**			
	1st most frequent	**2nd most frequent**	**3rd most frequent**	**Total % of the three most frequent realisations**
/a**ɪ**/	/i/	/ei/	/i:/	86%
/e**ɪ**/	/æ/	/ə/	/aɪ/	68%
/**əʊ**/	/**ɒ**/	/**ɔ**:/	/**aʊ**/	96%
/**aʊ**/	/**əʊ**/	/**ɔ**:/	-	100%
Average %				88%

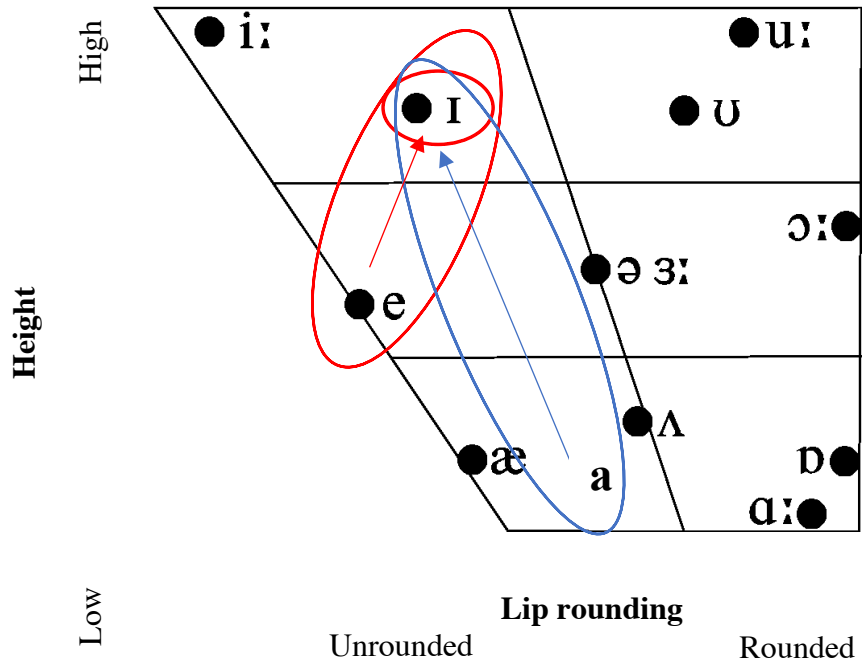


Figure 30. Standard RP vowel chart with notations showing the three most observed realisations (circled in red) of target diphthongl /ai/ (circled in blue).

Thus, to conclude the vowel section, error analysis of learners' decoding outputs highlighted difficulties with decoding English short vowels, long vowels, and diphthongs. Investigation of the nature of realisations produced for each vowel phoneme yielded a number of vowel realisations. Nonetheless, extensive analysis of these realisations suggests little evidence of randomness in learners' approach to vowel decoding.

4.2.1.2.Consonant-related errors

An error was classified as consonant related when the target consonant sound was substituted with an inaccurate one. As was shown earlier (Figure 11), consonant errors accounted for 22% of the errors observed in the whole sample analysis, of both the FPT and SPT datasets. Furthermore, when the decoding difficulties of each group were examined separately (using the SPT dataset), consonant-related errors emerged in all three proficiency groups with a largely similar proportional frequency (21%, 22%, and 19% of the errors observed in the A1, A2, and B1 groups, respectively). Given that consonant-related difficulties emerged in all three proficiency groups, the next step was to see if consonant error rates varied across the three proficiency groups. The data was first checked for normality, outliers, and homogeneity of variance using the same procedures outlined earlier. Results suggest that consonant error rates were normally distributed for the A2, $D(20) = .128$, $p = .200$, and B1 groups, $D(20) = .165$, $p = .159$. However, the consonant error rate of the A1 group was not

normally distributed, D(20) = .205, p = .027. Results from the Levene's test showed that these groups have homogeneity of variance, F(2,57)= 1.407, p = .253. No cases were identified as outliers.

As the normality assumption was violated, a non-parametric alternative of ANOVA, the Kruskal-Wallis test, was run. Results showed that there was a statistically significant difference in the consonant error rate between the different groups, $H(2) = 12.807$, p = .002, with a mean rank consonant error rate of 39.50, 31.98, and 20.02 for the A1, A2, and B1 groups, respectively. Pairwise comparisons with adjusted p-values (for multiple comparisons) showed that there were no significant differences between the consonant error rates of the A1 and A2 groups (H= 7.525, p = 0.511) and A2 and B1 groups (H= 11.950, p = .088). However, the difference between the A1 and B1 groups was statistically significant (H=19.475, p = .001).

In what follows, any problematic consonant phonemes found in the data are reported, along with the realisations produced for each. It was also of interest to calculate a facility level for these consonant phonemes using the same procedure outlined earlier. Table 18 below lists the facility level for every consonant phoneme coded in the phonological transcription data, and the following table (table 19) shows the realisations obtained for each consonant along with their percentage frequency.

Table 18. Facility level of consonant phonemes

Phoneme	Pseudoword item	N	Phoneme percentage of success	Real word item	N	Phoneme percentage of success	Average percentage of success
/p/	<pim>	60	46.67	<paper>	60	65.00	55.83
/b/	<barp>	60	76.67	<able>	60	81.67	79.17
/dʒ/	NA*	NA	NA	<jump>	60	58.33	NA
/g/	<ig>	60	71.67	<great>	60	100.00	85.83
/v/	<vasp>	59	64.41	<river>	58	87.72	76.06
/f/	<fet>	60	98.33	<fast>	60	100.00	99.17
/s/	<dess>	60	98.33	<stop>	60	100.00	99.17
/k/	<clirt>	56	64.81	<cat>	60	100.00	82.41
/z/	<zint>	49	93.88	<as>	60	86.67	90.27
/d/	<din>	60	96.67	<wood>	60	100.00	98.33
/tʃ/	<chur>	58	79.31	<child>	51	54.90	67.11
/ʃ/	<shlee>	56	92.86	<shoes>	59	98.31	95.58
Notes: *No pseudoword item was found for this phoneme. ** Items with N smaller than 60 were sampled from the FPT dataset.							

*** Three consonants were listed in an "other" category and are not included in this table. Consonant errors in this category were considered idiosyncratic (they were observed only once). These consonants are /r/, /m/, and /ð/.

Table 19. The realisations observed for consonant phonemes along with their percentage frequency

Consonant	Realisations observed and their percent frequency						Remarks
/p/	/b/ (96.9%)	/d/ (1.6%)	/t/ (0.4%)	/m/ (0.4%)	/k/ (0.4%)	/n/ (0.4%)	/t/, /m/, /k/, and /n/, appear to be idiosyncratic, as each was observed only once.
/b/	/p/ (80%)	/d/ (20%)					
/dʒ/	/ʒ/ (73.7%)	/g/ (26.3%)					•/dʒ/ was represented in the graphemes <j> (e.g, jump) and <g> (e.g., strange). • the realisation /ʒ/ was only observed in the <j> graphemic context. • /g/ was only observed in the <g> graphemic context.
/g/	/dʒ/ (88.5%)	/ʒ/ (11.5%)					
/v/	/f/ (95.7%)	/p/ (2.2%)	/w/ (2.2%)				• /p/ and /w/ appear to be idiosyncratic, as each was observed only once.
/f/	/v/ (63.6%)	/t/ (27.3%)	/l/ (9.1%)				• /l/ was only observed once.
/s/	/k/ (80%)	/z/ (20%)					•/s/ was represented in the graphemes <c> (e.g., space) and <s> (e.g., vasp) • /k/ was observed only in the <c> graphemic context. • /z/ was observed only in the <s> graphemic context.
/k/	/ʃ/ (4%)	/s/ (72%)	/sk/ (4%)	/g/ (20%)			•/k/ was represented in the graphemes <k> and <c>. • /s/ was observed only where <c> was the target grapheme, (e.g., clirt) • /z/ was observed only where <s> was the target grapheme, (e.g., vazp) • /ʃ/ and /sk/ appear to be idiosyncratic, as each was only observed once.
/z/	/s/ (100%)						• /s/ was observed only in the <s> graphemic context (e.g., is)
/d/	/b/ (75%)	/ð/ (25%)					• /ð/ appears to be idiosyncratic, as it was only observed once.
/tʃ/	/ʃ/ (100%)						
/ʃ/	/tʃ/ (100%)						

Two key points are worth noting from table 19 above: first, for consonant graphemes with only one phoneme correspondence, i.e., <p>, <b>, <d>, <ch>, <sh>, <f>, <v>, and <j>, the incorrect realisations reported often shared the same place of articulation of the target phoneme. For example, both /p/ and /b/ are bilabial sounds, /v/ and /f/ labio-dental, /tʃ/, /ʃ/ post-alveolar, and /dʒ/ and /ʒ/ are post alveolar. The only exception to this is <d>, as the most frequent realisation observed for this consonant was the bilabial /b/ (note also that for both <p> and <b>, alveolar /d/ emerged as one of the realisations). Second, for consonant graphemes with more than one phoneme correspondence, (i.e., /s/ and /z/ for <s>, /s/ and /k/ for <c>, and /g/, /ʒ/, or /dʒ/ for <g>), the incorrect realisations almost always pertained to the use of a regular but incorrect phoneme. For example, all errors decoding <c> in the pseudoword <clirt> pertained to the production of /s/ (for example, /slirt/), which is a regular phoneme for <c> in other contexts. Similarly, all errors decoding <g> in the pseudoword [<ig>, /ig/] pertained to the production of /ʒ/ or /dʒ/, (for example, /iʒ/), which are regular phonemes for <g>.

4.2.1.3.Phoneme insertion, omission, and mis-ordering errors

As the name suggests, this category comprises three types of errors: (a) phoneme insertion errors; (b) phoneme omission errors; and (c) phoneme mis-ordering errors. "Phoneme insertion" refers to cases where a participant adds one or more phoneme to the cluster of phonemes comprising the target item. For example, a reading of <play>, /plei/ as /pɪlei/ would be listed under this category, because the participant added the phoneme /ɪ/ to the target item. "Phoneme omission" refers to cases where a participant omits one or more phoneme from the cluster of phonemes comprising the target item. For example, a reading of <step>, /step/ as /tep/ would be included in this category because the phoneme <s> is omitted from the target item. "Phoneme mis-ordering" refers to cases where a participant mis-orders one or more of the phonemes comprising the item to be decoded. For example, a reading of <last>, /lɑ:st/ as /lɑ:ts/ would be included in this category because, whilst the reading comprises all the phonemes of the target item, it does not maintain the correct ordering of these phonemes.

In total, errors in this category accounted for approximately 17% and 12% of the decoding errors observed in the FPT and SPT datasets, respectively. They also emerged in all three groups with a largely similar proportional frequency (12%, 13%, and 11% of the errors in the A1, A2, and B1 groups). Having established that errors in this category emerged in all three groups of proficiency (note that all three sub-types of error within this category were observed in each group), the next step was to see whether the error rates varied across the three proficiency groups. The data was first checked for normality, outliers, and homogeneity of

variance using the same procedures outlined earlier. Results suggest that errors in this category were normally distributed for the A1, D(20) = .166, p = .148, and A2 groups, D(20) = .168, p = .139. However, the error rate of the B1 group was not normally distributed, D(20) = .300, p = .000. Results from the Levene's test showed that these groups have homogeneity of variance, F(2,57)= 2.908, p = .063. No cases were identified as outliers.

As the normality assumption was violated, a non-parametric alternative of ANOVA, the Kruskal-Wallis test, was run. Mean rank error rates were 36.02, 31.98, and 23.50 for the A1, A2, and B1 groups, respectively. Nonetheless, no statistically significant difference was found in the error rates between the different groups, $H(2) = 5.498$, p = .064 (though the p value was approaching significance). Upon closer examination, two cases – although not exceeding the minimum outlying threshold – appeared to be potential outliers with a z score of ~3 (particularly z= 2.53). When these two cases were removed from the analysis, a statistically significant difference was found between the error rates of the three proficiency groups, $H(2) = 9.161$, p = .010, with a mean rank error rate of 35.78, 31.82, and 19.94 for the A1, A2, and B1 groups, respectively. Pairwise comparisons with adjusted alpha levels (for multiple comparisons) showed that there were no significant differences between the error rates of the A1 and A2 groups (H= 3.950, p = 1.00) and A2 and B1 groups (H= 11.881, p = .084). However, the difference between the A1 and B1 groups was statistically significant (H=15.831, p = .010).

Furthermore, it is worth noting that when the realisations in this category where closely inspected, it was possible to identify particular situations in which phoneme insertion, omission, and misordering errors seemed to emerge.

For example, students appeared engage in **phoneme insertion** in three contexts: (1) the silent letter context, in which a learner adds to a word's target phonemes by way of pronouncing a silent letter, for example, decoding <knap> as /knap/; (2) two- and three- segment consonant clusters, in which a learner inserts a vowel to assist the pronunciation of the consonant cluster, for example, decoding <skree> as /sikri:/; (3) the double consonant context, in which a learner appears to geminate the consonant sound, for example, decoding <dess> as /desis/. However, it is worth noting that these contexts accounted for 83% of the instances of insertion and it was not possible to identify the patterns surrounding the remaining 17% of the observations. Yet, it is worth noting that some, though not all, of these other cases resulted in the production of a different real word, for example, reading the pseudoword <strone> as /strɒng/.

Additionally, when cases of **phoneme omission** where closely inspected, it appeared that students consistently engaged in phoneme omission in two contexts: (1) two- and three-

segment consonant clusters, in which a learner omits one of the consonants comprising the consonant cluster potentially to aid pronunciation, for example, decoding <skree> as /kri:/; (2) final <ee> context, in which learners appear to treat final <ee> as a silent letter, for example, decoding <shlee> as /ʃel/ (note that this error was coded in the phoneme insertion *and* omission sub categories, as the learner inserts a vowel to assist the decoding of the consonant cluster <shl> and also omits the final /i:/). Again, however, these two contexts accounted for 36.5% of the instances of omission, and no pattern could be discerned for the remaining 63.5% cases of omission. Nonetheless, a few of these 'other' omission cases resulted in the production of a real word, for example, reading the pseudoword <felly> as /fel/.

Finally, when cases of **phoneme mis-ordering** where further examined, one context was identified in which learners appear to consistently engage in phoneme misordering. This was the consonant cluster context, where learners appeared to mis-order phonemes so as to facilitate decoding of the consonant cluster, for example, decoding the pseudoword <shrattec> as /tʃærtik/ (this error was also placed under the consonant category given the incorrect decoding of /ʃ/). It is interesting to note here that, on several occasions, the consonant cluster was not part of the syllable structure of the target word, but nonetheless occurred as a result of learners' inaccurate decoding, more particularly, their articulation of a silent letter. To illustrate further, a decoding of <knap> as /kænp/ was listed in the phoneme misordering subcategory, as two phonemes, /n/ and /æ/, were misordered (it was also placed under the phoneme insertion subcategory given the insertion of /k/). In this example, the target <knap>, /næp/ is a CVC word *but* it is such only when it is decoded correctly. An attempt to pronounce the silent letter would result in the formation of the consonant cluster <kn> which does not exist in English or Arabic. In such situations, poor knowledge of silent letters altered the syllabic structure of the word (from CVC to CCVC) creating a further decoding challenge (i.e., not only that of silent letters, but also that of a consonant cluster) and resulting in the production of a realisation that deviates even more significantly from the target realisation (i.e., /kænp/ is more different from /næp/ than /knæp/). However, again it is important to note that the consonant cluster context accounted only for 32% of the instances of phoneme-misordering. It was not possible to identify the patterns surrounding the occurrence of the remaining 68% of the cases.

4.2.1.4.Whole-word errors

This category involves: (a) decoding errors which results in the production of a different real word (for example, decoding <able>, /eibəl/ as /<apple>, /æpəl/; and (b) errors which do not result in the production of a different real word but nonetheless deviated significantly from

the target realisation (for example, decoding <inside>, /insaid/ as /insɒl/. As noted earlier, there was some overlap between this category and other categories, because some realisations belonged to the "whole-word error" category and also to one of the categories discussed above (for example, a decoding of <paper>, /peipər/ as <pepper>, /pepər/ was placed under both the "vowel" category and the "whole-word error" category.

As was shown in Figure 11 earlier, whole-word errors accounted for 5% of the errors observed in the whole sample analysis, of both the FPT and SPT datasets. Furthermore, in a qualitative subsample analysis (Figure 12), whole-word errors emerged in all three groups of proficiency (A1, A2, and B1), with a relatively similar proportional frequency (7%, 3%, and 4% of the errors observed in the A1, A2, and B1 groups, respectively). Having established that whole-word errors emerged in all three groups of proficiency, the next step was to see if their frequency varies across the three proficiency groups. The data was checked for normality, outliers, and homogeneity of variance using the same procedures outlined earlier. Results suggest that whole-word error rates were not normally distributed for the A1, $D(20) = .218$, $p = .013$, A2, $D(20) = .290$, $p = .000$, and B1 groups, $D(20) = .278$, $p = .000$. No cases were identified as outliers and results from the Levene's test showed that these groups have homogeneity of variance, $F(2,57)= 1.270$, $p = .289$.

As the normality assumption was violated, a non-parametric alternative of ANOVA, the Kruskal-Wallis test, was run. Results showed that there was a statistically significant difference in the whole-word error rate between the different groups, $H(2) = 15.373$, $p = .000$, with a mean rank error rate of 42.38, 25.45, and 23.68 for the A1, A2, and B1 groups, respectively. Pairwise comparisons with adjusted p-values (for multiple comparisons) showed that there were significant differences between the whole-word error rates of the A1 and A2 groups ($H= 16.925$, $p = 0.004$) and A1 and B1 groups ($H= 18.700$, $p = .001$). However, the difference between the A2 and B1 groups was not statistically significant ($H=1.775$, $p = 1.000$).

4.2.2. Conclusion:

Error analysis of learners' decoding realisations highlighted difficulties with decoding English vowels and consonants, and a tendency to add to, omit, and mis-order words' phonemes, along with a tendency to make whole-word substitutions.

The series of between-group analyses suggest that no *qualitative* differences appear to exist in the nature of decoding difficulties, in that the above categories of error were observed in each group of proficiency. Nonetheless, *quantitative* variations were observed, as error rates in each of the emerging categories appeared to decrease as proficiency level increased. These

quantitative variations were often more apparent between the A1 and B1 groups than between the A1 and A2 or A2 and B1 groups.

Finally, in-depth analysis of the nature of the exact realisations observed in each category of error yielded several insights, the implications of which will be dealt with in the discussion chapter.

Having discussed the findings relating to the decoding realisations of Saudi foundation year ESL students, the next section presents the findings relating to their L2 decoding experiences.

4.3.The decoding experiences of Saudi foundation year ESL students

4.3.1. Results:

Analysis of participants' responses in relation to the decoding difficulties resulted in the emergence of six categories, explored in turn below. For organizational purposes, categories are presented in order of their frequency of occurrence. These are summarised in table 20 below:

Table 20. Summary of emerging categories

Categories
Lack of L2GPC knowledge (100%)
Word length (92%)
Unfamiliar words (75%)
Silent letters (50%)
English script (42%)
Other

4.3.1.1.Lack of L2 GPC knowledge

Lack of L2 GPC knowledge was the most frequent theme, occurring across all 12 interviews (i.e., in 100% of the sample).

Several responses related to general difficulties with the depth of the English writing system. For example, S3 noted "*in English, some letters are not pronounced as they should be"*. Similarly, S2 mentioned that she finds difficulty reading "*words which their pronunciation is different from the way they are written*". S6, too, noted that one issue that makes her decoding difficult is *"the word that has letters which, when together, they produce a certain sound"*, a view that was also shared by S9: *"when you merge two letters together like S and H and you get another letter [sound]"*.

Other responses were more focused as they related to difficulties with *vowel* GPC knowledge in particular, as S1 commented: *"The U. When is it /ei/ when is it /ju:/? I don't know. Sometimes even the O is not read as /o:/. Why do I not read it as /o:/? Like out of a sudden you have to read it as /ei/ or /ju:/"*. The exchange below with S5 indicates a similar difficulty:

S5: (points at the word <surrounded> as a difficult word)
Researcher: what made it difficult?
S5: the U, I don't know how to read it. I don't know how to read all these letters.
Researcher: which letters?
S5: the I and U and A and E.

In addition to vowel GPC knowledge, learners expressed difficulties with decoding double consonants. For example, S3 pointed at <worshipper> as a difficult to decode word and mentioned that the presence of the double consonants, <pp>, created a decoding challenge. Similarly, S10, pointed at the word <brass-work> and asked, "this [word]. Do I say this S twice? /bræsis/?"

S4 expressed difficulties decoding <ch> in <architecture>, as she commented, *"when it [<ch>] is like that I don't know if its /tʃ/ or /k/"*. S12 pointed at the same word, and following a few unsuccessful decoding attempts, she manages to decode <ch> successfully, but fails to correctly decode <-ture>, as the exchange below shows:

S12: (pointing at the word <architecture>) This. /a:rdʒ/. what? /a:rdʒtekt/. I don't know how to read it.
Researcher: what's difficult in it?
S12: (continues trying) /a:r/../a:rdʒ/.. Is this [arch] read tcha? Or.. just a second /a:rkitektur/?
Researcher: what was difficult?
S12: I always get confused when I see the <sh> and <ch>.
Researcher: its read /ɑːkɪtektʃə/
S12: aha so the <ture> you do it like that?
Researcher: yes tcha.

Decoding <c> was also problematic, as S10 noted, *"some letters I don't know how to pronounce…like this letter [<c>]. I see it C. But its not. Maybe its /ki:/. I don't know"*. S5 expressed a similar difficulty *"I don't know when to read it /s/ and when to read it /k/"*. S9 had a "funny incident" to tell regarding this particular difficulty:

S9: …Sometimes it changes the word. Sometimes you read the word and it changes completely. You want to deliver a certain meaning, but you deliver a different one.
Researcher: Repeat the last bit again
S9: This incidence has happened to me before. I wanted to read a word and instead of reading the C as /s/. Um this C should have been pronounced as [s] not [k].
Researcher: can you remember the word?
S9: Yes, it's a funny incident. I don't remember the word itself, but it means "ħayah" <حياة>[Arabic word for <life>], and I read it "ħimar" [Arabic word for <donkey>]. Do you now know it?

(The researcher thinks that the student has confused the words as <c> is not part of the graphemic structure of <life>)

Researcher: No. What's the word? can you not remember?
S9: Ok translate this for me <الرقص مع الحياة> /Araqsˁ maʕ alħaya/ (translates to "dancing with life" [Note that google translate would provide the translation "dance with life"].
Researcher: Aha "dance" and "donkey"

S9: Yes yes. I translated the word and then I read it as /k/ and it changed the whole word. The teacher asked us to say the last book we read. That's an Arabic book by Mahdi Mosawi. I translated the word dance. And then instead of reading the C as /s/ I read it as /k/. It became "donkey with life". So the teacher was like "are you sure this is the title of the book?". And I say yes. That's difficult. English. All of it is difficult.

Finally, before moving to the next category, a few points are worth adding in this section. First, on occasions, students appeared to have some knowledge of L2 GPC rules, nonetheless, that acquired knowledge was incorrect, as shown in the comment below made by S1 regarding the decoding of <tion>:

Researcher: and you said a difficult word here was "traditional. What did you say here?
S1: yes, the t, i, o, n you read all of it as sha. That's what I know, but sometimes.. see you just said "traditional". See you pronounced the N. How come? They told us the T I O N all of it is /ʃ/ and you pronounce the N? but why is it pronounced in this word?

Second, several comments, albeit suggestive of the existence of some L2 GPC knowledge, hinted at a lack of full internalisation of that knowledge, as shown in the example below:

S11: (The student pointed at<classifications> as a difficult word and tries to read it) classifications? That's the only pronunciation that I could come up with really.
Researcher: and why are you hesitant about this?
S11: I'll ask you first. Is my reading right or wrong?
Researcher: but why do you feel it might not be right?
S11: I don't know I just saw it and it scared me. I said maybe my reading is wrong.
Researcher: What's scary in it?
S11: maybe the T, I, O, N. I don't know.

Third, and finally, comments indicating lack of L2 GPC knowledge, were often accompanied by remarks suggesting the presence of negative affect (for example, "scared" "scary") (note the example above too):

S11: what is this word? (pointing at the word <Portuguese>)
Researcher: What's difficult in it?
S11: I don't know maybe the vowel letters.
Researcher: What's the problem with vowel letters?
S11: I don't know. They are many here. The word's shape is like scary.

4.3.1.2.Word Length

Word length was the second most frequent theme, occurring across 11 interviews (i.e., 92% of the sample). When learners were asked what exactly they found problematic in "long words", responses varied. The majority of the participants related their difficulties to the processing time needed to carry out one, or more, of the operations they engaged in to read

"long words" (for example, dividing a word into smaller parts, converting graphemes to their corresponding phonemes, and blending the phonological forms of these graphemes to generate the phonology of the whole word), as the comments below show:

S3: "I need time to read it well"
S5: "I need time to spell it. Divide it and like that."
S6: "I will take time till I am able to read it and spell it"
S8: "long words, it's just that even if I read the first part correctly, when I get to the second part, I would have to spell it, and you know you go back to the first part and then pull them together"

A few responses seemed to relate to changes in affective state which are triggered by the presence of long words. For example, S3 expressed how word length may discourage her from attempting to decode words:

Researcher: Here is a passage that you read before, which words would you say are difficult to read?
S3: this (pointing at the word <masterpiece>)
Researcher: what's the problem with it?
S3: I don't know up till now I don't know how to read it.
Researcher: what made it difficult?
S3: it could be the length that stopped me from trying in the first place.
Researcher: is any long word difficult for you?
S3: maybe. Sometimes it might not be difficult. But I just hesitate. I fear that if I read it my reading would not be correct, so I decide not to bother and read anyways.

Finally, it is important to mention that there was some overlap between this category and other categories, particularly the "lack of L2 GPC knowledge" and "unfamiliar words" categories. This is because, on occasions, follow-up discussions seemed to suggest that "long words" were mostly problematic when they were unfamiliar, and as such, learners would need to have a good command of L2 GPC knowledge in order to decode them, as shown below:

Exchange 1: (the participant previously mentions word length as one of her decoding difficulties)

Researcher: any easy words?
S3: <over>, <hold>, <can>
Researcher: why are they easy?
S3: you can count their letters. You can pronounce them quickly. You don't need to spell them. Or something like that.
Researcher: how about this word (The researcher shows her the word <passage>, a seemingly familiar word)
S3: (reads it correctly.) That one is fine.
Researcher: but it is not short if you compare it to <over> and <can>
S3: yes, but because we are used to it. Read it a lot.

Exchange 2: (the participant mentions word length as one of her decoding difficulties)

Researcher: Is any long word difficult to read for you?
S10: It depends on the letters. Like "shocolata" (Arabic word for chocolate which the student assumes is English too), yes it is long, but it is easy to read.

4.3.1.3.Unfamiliar words

Difficulties relating to unfamiliar words emerged in the interviews of 9 participants (75%). There appeared to be some overlap between this category and the "lack of L2 GPC knowledge" category. This is because difficulties emerging from 'unfamiliar words" were almost always related to learners' inability to retrieve the pronunciation of words from memory, thus necessitating the existence of L2GPC knowledge to facilitate real time computation of a word's phonological form.

Example 1:
Researcher: any easy words [in the passage]?
S2: Country names. <Casablanca>. (she reads it accurately). <Sign> (she reads it as /sing/. <centre> (she reads it correctly).
Researcher: why is <centre> easy?
S2: because I have seen it a lot.
Researcher: is any word you have previously seen easy to read?
S2: In the beginning it would be difficult, but once I memorise its pronunciation it would be easy…
Researcher: and when you dont manage to memorise the pronunciation?
S2: it would be difficult to read.

Example 2:
S1: When you read words you don't know…there are many words that I know. Many words. I understand their meanings very well. You speak with me in English, I could understand and all, but the problem is when I myself begin to read, I look and say what is this word. I don't understand what this word is. but when I ask someone to read this word for me, the moment they read it, I automatically recognize the word and say ah.
Researcher: can you give me an example
S1: like <impossible>. The first time I read it I was like /im/../pu:/../si:/../bul/. I did not know how to read it, but when I asked someone to read it, I said ah impossible, I know this word. I know the meaning of this word.
Researcher: and how did you learn this word impossible?
S1: I don't like to learn a thing like learn from a book and like that. I watch American movies and series.

While the above example appeared to be related to familiarity with the *written* form of the word, it is interesting to note that some responses appeared to be related to familiarity with the *spoken* form of the word. For example, S7 was hesitant about the pronunciation of the word

<distinguish>:*"/distingiʃ/? I don't know something like that*". When asked why she was hesitant, she replied, "because I have not heard it. I like to hear the word before I say it". The researcher then asked her "what if you tried?". And she responded, "*It won't be useful. It's a new word. even if it was short. I have to hear it first to read it correctly. I don't like to read words incorrectly*". Similarly, S6 mentioned that she encounters difficulties reading long words. When she was asked to show the researcher examples of words she finds easy to read, she mentioned the word <agreement>. The researcher then asked her why she finds the word easy given that she previously mentioned difficulties with long words, and the student commented, "*I heard it a lot*". She was then asked if words that she has not previously encountered would be difficult, and she commented: "*it depends on the letters and sounds if I can produce them. But if its long and it's the first time I see it, it would definitely be difficult. If I have not heard how its read, for sure, I won't know how to read it"*. A further example is provided below:

> *Researcher: any easy to read words [in the passage]?*
> *S4: <kings>, <them>, <from>, <begin> (she reads them accurately)*
> *Researcher: why are they easy?*
> *S4: they are short, and I know them previously.*
> *Researcher: is any word that you know easy to read for you?*
> *S4: yes*
> *Researcher: any word that you have not seen before is difficult to read?*
> *S4: yes, because I might not have heard it before. When I see it for the first time, I would not know how its pronunciation should be.*

Finally, it is worth mentioning that some participants noted that unfamiliar words would not be problematic if they were "short". For example, S7 mentioned that "*New and long words are difficult*". When she was asked if this is the case with any new word, she replied, "*No, if it was short it would be easy"*. Similarly, S12 noted that new words would be difficult to read: *"It depends if I know the word, because I am not that good in English*". When she was later asked if this is the case with any new word, she replied *"it depends. I would need to see it. If it's long, it would be difficult because I would need to divide it for me to be able to read it".* Similarly, when S8 was asked the same question, she replied, "*Not always. If I don't know it and its short, I would be able to read it".*

4.3.1.4.Silent letters

Difficulties relating to silent letters emerged in the responses of 50% of the sample (N=6). Some of the responses in this category reflected general difficulties with English silent

letters, i.e., difficulties arising from lack of awareness of what, and when, letters might function as silent, as S2 commented: "*What I have problems with most is silent letters. I did not memorise what letters are silent, and I end up reading them".*

Some responses were more focused, reflecting difficulties with silent letters in *unfamiliar* words particularly, as S11 noted: *"Sometimes, I read a word correctly, but what confuses me is .. If the word has a silent letter. Especially if it was in the middle. I collapse. I feel my pronunciation is wrong. Mostly, I won't know how to read it unless I know the word. But a new word and has a silent letter. No, I won't know*". S1, too, made a similar comment: *"Ones [words] that have letters that are not pronounced. When you read words you don't know. Do I read these letters with the word or do I ignore them?"*

4.3.1.5.English Script

When asked how they find the English script, 5 participants (42% of the sample) expressed difficulties which appeared to be related to the visual characteristics of the English script. For example, S1 commented,

> *English letters are all similar. In Arabic its different like the qaf [Arabic consonant letter representing the sound /q/. No English equivalent exists for this letter/sound] is different from whatever. Each letter has its own shape. OK, in English they do have letters that have unique shapes. But when the letters are all small small small you find them all small and mixed and similar. Like the E and A and U and we and F, but no the F and T are slightly bigger you can distinguish them, but the rest they are small. I don't know.*

Though this view was not shared by all students, as S3 comments below (note that, nonetheless, she expressed *past* difficulties with English script):

> *I find the shape of English letters simpler than that of Arabic letters. The letters do not have complexity in them. Like even if you link the letters together when writing the letters will still look the same unlike in Arabic when you link the letters their shapes change. But in the past I used to find the g and y, when written as small letters, they used to look similar. The d and b, until very recently, I used to confuse them which is to the right which is to the left.*

Several letters emerged as visually problematic in discussions of English script. For example, S5 commented, "*The q and p. when they are small [They confuse me]".* The three exchanges below highlight difficulties differentiating <t> and <f>, <I> and <l>, and <p> and <b>, respectively:

- **<t> and <f>**

S2: (tries to read <traditional>). Its pronunciation is just difficult. /træ/.. /di:/../fi:/../ʃinəl/. Is this right?
Researcher: but I don't see a fa here.
S2: Ah that's a T. traditional. Traditional (reads it correctly).
Researcher: why did you say f?
S2: because the t looks like F

- **<L> and <I>**

S9: (pointing at the word <Islamic> as a difficult word and tries to read it; note that the font used in the passage presented <I> without the upper and lower strokes <Islamic>)
S9: /læs/.. /læsla/. What is that L S L.
Researcher: that's not L its I. Capital I.
S9: but they are written similarly I and L. If you just add a mark showing that this is an I. You know that the I when its capital there are those two strokes up and below. Just add them.

- ***<p> and <b>***

Researcher: how do you find the shapes of English letters?
S9: the B, those that are written like this the stick down and the circle up, or the stick up and the circle down, are they all B or what?
Researcher: one is B and that has the stick above and one is P and that has the stick below
S9: So there is a difference

4.3.1.6.Other

A smaller number of participants referred to decoding difficulties other than the ones described above (N=4). For example, two responses were coded in a **"phoneme articulation**" category. This category was formed to account for responses which appeared to be purely related to difficulties *articulating* L2 sounds, as opposed to general difficulties with L2 GPC knowledge. For example, S2 commented: *"The <ch> which you read as tcha… I feel it's difficult for my tongue to say it*". In this example, the student clearly knows the corresponding phonological form of the grapheme <ch>; nonetheless, she encounters difficulties articulating that form/phoneme. Similarly, S6 mentioned, "*The vowel sounds. I just don't know how to produce these sounds*" (though it is possible that the student was referring to a lack of vowel GPC knowledge, too, in this example).

Furthermore, one response was coded in a **"tongue-twister"** category. This category was formed to account for responses which seem to be related to difficulties articulating repetitive sounds, as the following example shows:

S4:the letters like in /stætis/.. /stætis/ (the student encounters difficulties saying the word <statistics>). The word /eħsˤaʔ/ in English (/eħsˤaʔ/ is the Arabic equivalent of <statistics>). That is very difficult.

Researcher: Whats difficult in it?
S4: the letters are repeated more than once

Finally, one response was given the code **"prior knowledge of a graphemically-similar word".** This code reflected a situation in which a learner's prior knowledge of a word, that is graphemically similar to the target word, contributed to an inaccurate reading of the target item (possibly through a mechanism in which the phonological form of the graphemically-similar word competes for selection, potentially encouraging partial/minimal processing of the target item), as illustrated below:

Researcher: show me an easy one?
S5: /pleis/ (the students points at the word <palace> and reads it as <place>)
Researcher: why is it easy?
S5: I know it. I read it a number of times.
Researcher: is it the same word? the word that you read numerous times?
S5: (looks at the word again): yes.
Researcher: look at it carefully
S5: no. palace. (the student reads it correctly)
Researcher: why did you say place initially
S5: but I just looked at it and thought it's the word place.

4.3.2. Conclusion

To conclude this section, thematic analysis of the interview data highlighted a range of decoding difficulties. These related to lack of L2 GPC knowledge, difficulties reading long words, difficulties reading unfamiliar words, lack of silent letter knowledge, and lack of familiarity with the English script. Having presented the findings relating to the decoding experiences of Saudi foundation year ESL students, the next section presents the findings relating to their L2 decoding processes.

4.4.The decoding processes of Saudi foundation year ESL students

4.4.1. Results:

Analysis of participants' responses in relation to their decoding processes resulted in the emergence of the following seven categories: (1) automatic processing; (2) small-unit processing; (3) large-unit processing; (4) knowledge-informed processing; (5) analogy-guided processing; (6) intuition-guided processing; and (7) a further "other" category which was primarily formed to account for the few instances of "partial processing" observed.

For organizational and presentational purposes only, these categories are presented under three headings below:

(A) Automatic processing (category 1 above);

(B) Assembled processing (which included categories 2, 3, 4, 5, and 6 above); and

(C) The "other" category.

4.4.1.1.Automatic processing

A response was given this code if it reflected lack of engagement in any thinking prior to decoding words. This lack of engagement was evident in responses along the lines of "I just said it", "I know it", "I've memorized it previously", etc. (Woore, 2010). Comments of this nature constituted 28.6% of the total comments produced.

A sample of the responses coded in this category are provided below:

- *S1: <him> "/him/. Normal… I didn't have any thought"*
- *S2: <no> "/nou/… Like its something I am used to, so normal"*
- *S4: <jump> "/ʒʌmp/… no thoughts"*
- *S5: <remember> "no thought came [to my mind]"*
- *S6: <find> "/find/ that's a word I already know"*
- *S9 <him> "/him/ I know it. I mean I've learned it before"*
- *S11: <great>: "/greit/ again its clear"*

When responses in this category were closely inspected, four observations were noted. First, the majority (81%) of the responses in this category were produced when learners were presented with real word stimuli. In other words, from a descriptive standpoint, learners were more likely to engage in automatic processing when reading real words than when reading pseudowords. Although it is unclear how a pseudoword would be decoded completely automatically, it is possible that some level of conscious computation was involved when

generating the phonological form, but that the computation process was nonetheless too quick to be consciously accessible and reported on.

Second, instances of automatic processing were not always associated with the production of accurate outputs. For example, when presented with the word <plane>, S15 quickly produced the realisation /plæn/ and commented "I already know it". The researcher then followed up by commenting "and how about that <e> at the end", to which she responded, "it is silent" (potentially suggesting that (a) the learner had noticed the vowel grapheme <e> but, nonetheless, assumed that it does not play any role in altering the pronunciation of the preceding vowel; or (b) that the previously-acquired phonological form for that item was incorrect). Similarly, S4 decoded <j> in <jump> as /ʒ/, /ʒʌmp/, and expressed that she did not engage in any thinking prior to her decoding.

Third, automatic processing was not necessarily employed at the whole word level, as there were occasions when learners appeared to engage in automatic processing at units smaller than the whole word (for example, syllable level). For instance, when presented with the item <garment>, S8, appeared to decode the first syllable <gar-> letter by letter but mentioned that she read the second syllable, <-ment>, *"altogether*". Similarly, when S14 was presented with the monosyllabic pseudoword <knap>, she commented *"/knæp/ I know it the word <nap>, I just added k to it*", suggesting that automatic processing had possibly taken place when decoding the <nap> segment.

Fourth, though not of primary interest here, it is worth reporting that some accidental comments seemed to suggest that automatic but *inaccurate* decoding could potentially interfere with meaning access. For example, S14 mentioned that she had no thoughts when she was presented with the pseudoword <prain>, which she quickly, *but incorrectly*, decoded as /brein/. Nonetheless, she further added *"and this word means <aqil>?"* (Arabic equivalent for <brain> (though it is possible here that this confusion may have merely, or also, arisen from orthographic similarity between <brain> and <prain>). However, it is also worth noting that there were occasions when automatic *but inaccurate* decoding did not appear to interfere with meaning access, as the exchange below indicates:

(presented with the item <plane>)

S8: /plæn/. Normal (i.e., I did not have any thoughts)
Researcher: and what does this word mean?
S8: <tʔayyara> (Arabic equivalent for <plane>)
Researcher: Is this how we read it?
S8: yes

Finally, although responses in this category seemed to reflect automatic processing, it is difficult to eliminate the possibility of learners producing these comments as a result of difficulties verbalising their thoughts. That is, learners may also say "I know it", "I just said it" because they were unable to articulate their thinking. Unfortunately, the think aloud methodology does not allow us to test this possibility.

4.4.1.2.Assembled processing

Categories grouped under this heading were indicative of learners' attempts to "assemble" (or compute) the phonology of the word by means of:

(1) dividing words into units, small and/or large (herein referred to as *small-unit processing and large-unit processing*, respectively), and/or by:

(2) conscious implementation of L2 GPC rules. As will be discussed later, these L2 GPC rules were *mostly* informed by (a) prior abstract symbol-sound correspondence knowledge (referred to as *knowledge-informed/mediated processing*) and/or by (b) analogy-based reasoning (referred to as *analogy-guided processing*). It was mostly *and not always* informed by these two because, as will be shown later, on occasions, learners seemed to appeal to their "intuition" when converting graphemes to their corresponding phonemes (referred to as *intuition guided-processing*).

Responses reflecting assembled processing accounted for 70% of the total responses observed. To ensure thorough reporting of data, each of these categories is dealt with individually in this section.

4.4.1.2.1. Dividing words into units

- Small unit processing

Small unit processing was observed when learners attempted to divide a word into small units, often at the single grapheme level. Examples of responses reflecting learners' engagement in small-unit processing are provided below:

- *S1, <om>: "/ɒm/..Here I just attached the O to the M"*
- *S6 <garment> "[I read it] following the letter ordering"*
- *S7 <hedfert> "I sounded it out it letter by letter"*
- *S11: <frip> "I pronounced each letter and then pulled them all together"*
- *S13 <guddy> "this is a new word so I pronounced each letter"*
- *S15: <lat> "/læt/ you've got L and then A and then T"*
- *S15: <frip> "here I went one letter after another. F and then R and then I and P"*

It is worth noting here that, on occasions, this small unit processing was associated with the gerneration of an incorrect decoding output. This was often – though not always – the case when learners attempted to divide a single grapheme, comprising of two or more letters (for example, <sh>, corresponding to the single phoneme/ʃ) into two or more graphemes (for example, <s> plus <h>), and consequently, phonemes (for example, /sh/). In the present data, this practice was often observed when decoding double consonant graphemes (for example, <dd> in <muddy>), the <kn>, <ai>, and <a-e> graphemes.

To illustrate, when presented with the pseudoword <guddy>, S15 commented "*Here is G U and then its like you pronounce the D twice. Like that /guddi/"* (the participant geminates the /d/ sound). Similarly, when presented with the pseudoword <shrattec>, several participants appeared to engage in small-unit processing when decoding the double consonant grapheme <tt> (or at least express difficulties decoding it), as shown in the responses below:

- *S1: "ʃrætitis/ the S with the H becomes sha. /ʃrætitis/… No wait… /ʃreitis/, but if so, I would be reading one T only (asking the researcher)"*
- *S8: "I spelled this and then said the first part is shra… This double t is causing some trouble."*
- *S9: "what's that… That's weird to be honest... Like the C is /k/… and the double T means that you make the T longer /tt/ (pronounces the /t/ sound geminated).*
  Researcher: when you have the same letter repeated you make it longer?
  S9: yes
- *S11: <shrattec> /ʃrætis/ the S and H becomes sha so it becomes /ʃrætis/. the T maybe we make it heavier because it's double.. (the participant tries to geminate the T)… No, we won't double it...*

Furthermore, when presented with the item <knap>, S1 appeared to treat the <kn> grapheme as two graphemes, <k> and <n>, and consequently converted them to two phonemes /k/ and /n/, commenting: "/knæp/ I attached the letters together. Letter after another". Similarly, when presented with the item <prain>, S11 commented "/prain/ "P R A I N", possibly suggesting that the learner has divided the grapheme <ai> into two graphemes <a> and <i> corresponding it to two monothongs /a/ and /i/, as opposed to the diphthong /ei/.

Similarly, small unit processing was problematic when it was employed to decode the split diagraph <a-e> (corresponding to the /ei/ sound). This was the case because learners appeared to treat the <a-e> grapheme as two individual graphemes <a> and <e>. On occasions, the two graphemes <a> and <e> were converted to the phonemes /æ/ and /i/ or /i:/, respectively. For example, when presented with the word <waves>, S5 mentioned: <waves>: *"/wæviz/ W and A*

and V and E and S and then I read them together". In a similar manner, when S3 was presented with the pseudoword <depate> she commented: "/di:pæti:/ here I just linked the letters together. One after another". On other occasions, the two graphemes were also treated as individual graphemes, but while the <a> was converted to the phoneme /æ/, the <e> was treated as a silent letter that plays no role in altering the pronunciation of the preceding <a>. For example, when presented with pseudoword <pave>, S5 commented: /bæv/ "B and A and V and the E you don't pronounce it". Similarly, when presented with the word <plane>, S6 commented: "I went according to the letter ordering and then said, the E at the end, you normally don't pronounce it". Thus, in all of these examples, it is clear that learners have processed the vowel graphemes contained within the <a-e> cluster, yet, appeared to employ a small-unit processing mechanism when decoding these vowels.

- Large unit processing

Large unit processing was observed when learners attempted to divide a word into large units, often at the syllable level. For example, the responses below appeared to reflect learners' engagement in large-unit processing:

- *S2<depate> "/dipeit/ we divide this into <de> and then <pate> and the E at the end I did not feel its pronounced"*
- *S1: <hedfert> "/hedfert/ until here [pointing at d] that's /hed/ and then /fert/"*
- *S3 <garment> "/gærmint/ I divided this into <gar> and <ment>*
- *S3 <remember> "/rimembir/ I divided it into <re> <mem> and <ber>"*
- *S9: <fermabalt> "/ferma/ alone and then /balt/ alone."*
- *S10: <hedfert>. "/hedfert/ This we divided it into two. So it's one word but we divide it into two"*

When comments reflecting large unit-processing were closely inspected, two observations were noted:

- The first observation is that, on occasions, large-unit processing was associated with the production of an incorrect decoding output. This was the case when learners attempted to incorrectly break up words inappropriately, suggesting a lack of English syllable division knowledge (for example, segmenting monosyllabic words, or inadequate segmentation of di-/multi- syllabic words). Relevant examples are provided below:

Breaking up monosyllabic words

- *S1: <waves> "/wʌfis/ I think this is how you read this word. The W with A sounds like /wæ/ and then you read the /fis/ together"*
- *S1: <pave> /beivi:/ "here I split the word into /bei/ and /vi:/. So its like two words in one"*
- *S1: <prain> "/prein/../prain/.. /preiʔin/. Here there is /prei/ plus /in/"*
- *S3 <strone> /stroʊn/... /stroʊni:/ "I Divided it into <stro> and <ne>"*

Incorrect (non-canonical) segmentation of bi-/multi-syllabic items

- *S3 <pleasant> /pli:sænt/.. /pleisænt/ I divided it into <plea> and <sant>*
- *S5: <budget> "/bʌdget/ I divided this word twice [into two] bud and get"*
- *S12<depate> "/di:pʔeit/. This word needed to be divided deep and ate. Even when you memorise it, you would need to divide it like that."*
- *S3 <limousine> "/li:muju:si:n/ "I divided this into "<limo> /limu/" and "<usine> /ju:si:n/"*

However, it should be noted here that incorrect division was not *always* associated with the production of incorrect outputs. For example, when presented with the item <depate> (see exchange below), S11 appeared to inadequately segment the item into <dep> and <ate>, (instead of <de> and <pate>), nonetheless, the final output that she produced was accurate (i.e., no glottal stop was present after the /dip/ segment). However, these occasions were the exception, as incorrect division was almost always associated with the generation of incorrect decoding outputs.

S11: I divided it into two parts <dep> and <ate>.
Researcher: so the second part where does it start?
S11: where the A is
Researcher: can you read the first part
S11: /dip/
Researcher: and the second
S11: /eit/
Researcher: and altogether what's the word?
S11: /dipeit/ (the participant decodes the item accurately)

- The second observation was that correct division was *not* necessarily associated with the generation of accurate decoding outputs. This was clear when:
  (a) learners were able to segment words successfully but appeared to lack accurate L2 GPC knowledge. For example, S13 was able to successfully break <garment> into <gar> and <ment>: *"/dʒarmint/. I divided it into two parts so I can read it. /dʒar/ and /mint/*". Nonetheless, the decoded output was incorrect, as the learner assumed the <g> grapheme in this context would correspond to the phoneme /dʒ/. Similarly, S4 divided <guddy> into <gu> and <ddy>: "/ge:di/. I divided this into /gei/ and /di/". The following exchange then took place:

  Researcher: what is the letter after G?
  S4: U
  Researcher: and how is it read?
  S4: sometimes you read it as /ei/

  (b) learners were able to segment words successfully but appeared to encounter difficulties *articulating* L2 phonemes. For example, describing her thought processes, S4 appeared to be able to segment <depate> successfully: *"/dibæt/ I break this into <de> and <pate>"*. Nonetheless, the learner appears to be unable to articulate the /p/ sound), as the following exchange suggests (though note that in this example, the participant also lacks L2 GPC knowledge, as evident in her incorrect decoding of the <pate> segment):

  Researcher: what is the third letter
  S4: B
  Researcher: this is a B?
  S4: yes
  Researcher: how about the P (the researcher clearly articulates the /p/ sound)?
  S4: yeah it's a B.
  Researcher: B or P?
  S4: B the strong one (the student means P but cannot articulate the sound). I pronounce it /b/.

  (c) learners were able to segment words successfully but nonetheless encountered difficulties blending (pulling together) the phonological forms of the correctly segmented parts to produce the final whole-word phonological form. For example, when presented with the word <intuition>, S1 commented "I can't read it". When the researcher asked her to describe her thoughts processes, she appeared to have somewhat managed to segment the word into syllables: "here is <in> /in/ and then

this <tui> /tju:/ and this word <tion> can't be divided". Nonetheless, she was unable to pull the parts together to produce the phonological form of the whole word: *"I can't read it"*.

4.4.1.2.2. Conscious use of L2 GPC rules

Responses in this assembled processing category often reflected conscious solving of L2 GPC "problems". The computation process was mostly informed by: (a) prior abstract symbol-sound correspondence knowledge (68.6% of the time); and/or (b) analogy-based reasoning (24.8% of the time). However, as mentioned previously, there were occasions when learners seemed to appeal to their "intuition" when converting graphemes to their corresponding phonemes (6.6% of the time). In what follows, results relating to each of these approaches to grapheme-phoneme conversion are presented.

- Referencing prior abstract L2 GPC knowledge

Abstract GPC knowledge here refers to the presence of prior knowledge concerning the correspondence between a given grapheme and phoneme (for example, knowledge that grapheme <l> corresponds to the phoneme /l/). On occasion, learners' prior GPC knowledge was correct, as the following examples show:

- S11 <waves> *"/weivz/. Here it all depended on this vowel <e>. The W and then A that's a vowel. The V is not a vowel. Then there is the E so you lengthen the A and shorten the E and it becomes like that /weivz/"*.

- S5: <knap> *"/næp/ I said, the K, if there is N before [after] it, then K should be silent"*

- S12: <strone> *"we wouldn't say /ɒ/ because there is an E and I think it would influence the pronunciation of the <o> so it becomes /stroʊn/. I did not pronounce it like <om> I made the <o> clearer like its lengthened because of the E which I think affects how you should pronounce it"*

- S15: <shrattec> *"I pronounced this <sh> sha.. sha not tcha.. sha because its <sh>. And I divided it into <shra> and <ttec>"*

However, on others, that prior knowledge was inaccurate, as shown in the example responses below (though note that on occasions, that knowledge was "partially correct", in that applying it in other contexts could yield correct responses; see the first example below). For organizational purposes, examples reflecting similar inaccuracies are grouped together:

- **/e/ in <ea>**
  - *S2 <pleasant> "/pli:zænt/ this is how you read the E and A when they are together. They have a particular pronunciation. And then <sant>. Since there is A it is read like that /sænt/.*

- *S7: <pleasant> "/pli:zænt/) I read this letter by letter and the E and A when together they become /i:/"*

- **/i/ in <I>**
  - *S3 <him> "/him../haim/. /him/ no /haim/ there is I".*
    Researcher: why?
    S3: its I not E.

  - *S12: <frip> "/fraip/" here I said there is no E at the end of the word so it wont influence the way I pronounce the I so you pronounce it /ai/ /fraip/. This is obvious."*

- **/ei/ in <a-e>**
  - *S3 <plane> "/pli:n/ there is the E at the end so you make this sound /li:n/"*
  - *S4: <waves> "/wævz/ the E normally, I have this bit of info, but I am not sure it's right or not, but its that the E is often not pronounced"*

- **Silent letters**
  - *S3 <budget> "/budit/ I first read the part at the beginning <bud> and then <et>"*
    Researcher: and the G?
    S3: silent

  - *S6: <fermabalt> /fremabalt/ two parts. Frema and balt.*
    Researcher: whats the first part?
    S6: frema. From the F until the A and then <balt>.
    Researcher: whats the letter after the f?
    S6: e (student reads the word correctly)
    Researcher: ferma or frema?
    S6: if the E is silent then it would come like frema.
    (Note: In the above example, the incorrect knowledge concerning the silent function of <e> seemed to have contributed to the formation of the consonant cluster <frm>, which in turn contributed to the insertion of the vowel <e> /frem/ to aid pronunciation of that cluster)

- **/ei/ in <ea>**
  - S4: <great> "/gri:t/ you read it /gri:t/ the first letter is E and then there is the A"
- **/ʌ/ in <u>**
  - *S4: <guddy> (decodes it as /ge:di/)"I divided this into /gei/ and /di/"*
    Researcher: what is the letter after G?
    S4: U
    Researcher: and how is it read?
    S4: sometimes you read it as /ei/

  - *S6: <budget> "/budit/.. I thought the G would be silent and the rest I went according to the letter ordering"*

- **// in <o-e>**
  - *S5: <strone> /strɒn/ again the E is not pronounced so it becomes /strɒn/*
- **/g/ in <g>**
  - *S6: <guddy> "/gudi/ or /ʤudi/? I feel it would be /ʤudi/ if the letter is J not G.*
  - *S12: <garment> "/ʤermʌnit/" this G is a little different. We would not pronounce it /g/ no its /ʤ/ /ʤermʌnit/... But maybe its pronounced /g/ /germʌnit/. No I don't think so. I often notice that whenever the <g> is at the beginning of the word its pronounced /ʤ/.*

- **/k/ in <c>**
  - *S11: <shrattec> /ʃrætis/ the S and H becomes sha so it becomes /ʃrætis/. the T maybe we make it heavier because it's double… (tries to read it while making it double). No we won't double it. The C we make it sa or ka? Si probably. Yeah highly likely its si. So it becomes /ʃrætis/.*
    Researcher: and why c?
    S11: because there is no letter after it so that we pronounce it ka. So it becomes /ʃrætis/.

  - *S10: <shrattec> "/ʃærtik/ no not /ʃærtik/. /ʃrætik/. It can be /ʃrætik/ and it can be /ʃrætis/. It depends if you are going to pronounce the [C as] si or k.. but its /ʃrætis/.. Yes, yes its /ʃrætis/. Because the C is pronounced k if its followed by vowel letters. Yeah I know. /ʃrætis/. So like <car> after it is <a> so its pronounced k /ka:r/.*

  - *S12: <calculated> I don't know there is a C here. I can't distinguish between… the C when is it pronounced /s/ and when it is pronounced /k/ so maybe I'll get confused. Maybe the first time it's /s/ and the second time it's /kk… or the other way around… or both /s/ … or both /k/. I don't think I was able to read it. Maybe if I divide it I would know?*
    Researcher: how would you divide it?
    S12: I would divide it into three bits.
    Researcher: could you give it a go?
    S12: if we say both are /k/ it would be /kalkilited/ or /si/ and it would be /salsilited/

- Analogy-based problem solving

In the field of decoding, analogy-guided processing refers to using the pronunciation of a familiar word to support the decoding of an unknown one which shares graphemic similarity with the familiar word (Woore, 2010). As detailed in Woore (2010), this process comprises three stages. Upon encountering a word containing a difficult grapheme (say the <kn> grapheme in <knock>), learners will first identify a word (the analog) that shares graphemic similarity with that difficult word (for example, <<u>kn</u>ow>) and retrieve the pronunciation of that word/analog (example, /nəʊ/). Second, learners will extract the pronunciation of the "difficult" grapheme(s) from the pronunciation of the whole word (analog) (for example, <kn> → /n/); Third, learners will integrate the pronunciation of that segment into the pronunciation of the

other graphemes in that word to produce the final phonological form of the whole word (for example, /nɒk/. Unsuccessful execution of any of these sub-operations could contribute to the production of an inaccurate decoding output (ibid).

On occasion, this analogy-guided processing appeared to contribute to accurate decoding of the problematic grapheme, as the following examples show:

- S2: <pave> *"/peiv/ I thought this word is like <game> and in it the E is not pronounced"*
- S7: <knap> *"this /næp/ is like <knife> I think the K will not be pronounced"*
- S7: <prain> *"/prein/ this looked like <brain> but the B is P here"*
- S9: <pave> *"/beiv/ this one is like <brave> but without the R"*
- S10:<depate> *this is somewhat like the word 'update' but the letters are different. No no no. It is so much different. But the ending is the same.*

However, there were occasions when analogy-guided processing was associated with the production of inaccurate outputs. This was often the case when learners failed to perform successfully one or more of the sub-operations mentioned earlier. Particularly, nearly all unsuccessful analogies found in the data arose from difficulties identifying a word that shares graphemic similarity with the target item (stage 1 above). For example, when presented with the item <budget>, S12 expressed that she encountered difficulties decoding the item: "*bɒdet/ or /bɒdegt/... this word I didn't know how to pronounce it*". She then expressed that she found the <g> grapheme particularly difficult, as it could function as a silent letter, as in the presumed analog <bought>: "*sometimes the G becomes silent like the word <bought> or like that. I sort of did not how to pronounce it*". Clearly here, by selecting the analog <bought>, the student failed to identify a word that shares graphemic similarity with the target item, as <budget> does not contain the <gh> grapheme present in <bought>. Similarly, when S13 was presented with the item <strone>, she produced the output /strɒn/ and commented "/strɒn/ this is close to which is why I knew how to pronounce it". Again, the student here appears to have failed at identifying a word that shares the *necessary* graphemic similarity with the target item (i.e, the <o-e> cluster, though the analog , may have contributed to the accurate production of the <str> segment in <strone>)

Furthermore, it is interesting to note that there were instances when students managed to execute each of the sub-operations noted earlier successfully, but nonetheless, produced an incorrect decoding output. This was the case when (1) learners appeared to draw analogies to "exception" words. For example, when presented with the item <pave>, S6 commented "/bæv/ I thought this looks similar to the word <have> but there is the B instead of the H". Similarly,

when S12 was presented with the word <pleasant>, she produced the output /pleisænt/ and commented "the E and A when together I thought they would be pronounced like in the word <great>". In both examples, the pronunciation of <have> and <great> are considered exceptions; (2) learners appeared to draw analogies to the pronunciation of the stem of the word, potentially unaware of the role prefixes and/or suffixes may play to alter the pronunciation of that word as the two examples show:

- *S6: <pleasant> /pli:sent/ "according to the letter ordering and the A and E its like the word please you read it and then I just added /ænt/*

- *S10: <pleasant> /pli:zent/ this is derived from <please> or maybe <pleased> like <happiness> like "I am pleased". It's the same but the word ending is different so you read it /pli:zent/*

Finally, it was mentioned in the beginning of this section that, on occasion, learners appeared to be consciously trying to compute the phonology of the words, nonetheless, that computation appeared to be neither informed by prior L2 sound-symbol knowledge, nor by the use of analogies. Instead, learners seemed to appeal to their "intuition" during the computation process. Again, at times, that intuition was successful, as the examples below suggest:

- *S1: <Guddy> (student appears to rely on intuition when decoding the <g> grapheme) "/gudeidi/ But is this pronounced /dʒi/ or /gi/. I mean /gudeidi/ or /dʒudeidi/ so I don't know which one exactly, so I'll have to just choose (student repeats both readings to choose) /gudeidi/"*

- *S9: <garment> /gærmint/no I can't say /ʒærmint/ here no its /g/.*
  Researcher: why?
  S9: /g/. the word is pronounced with /g/. Like there are words you can say /ʒ/ and words you would say /g/.
  Researcher: and how could one know?
  S9: No I don't know. I go with what suits the word. Like I read it and decide if its /g/ or /ʒ/.

- *S10: <frip> /fraip/… /frip/ its either /fraip/ or /frip/, both words would be suitable.*
  Researcher: do you mean both are correct?
  S10: Yes. No. It's one word that is correct but I can read it in two ways.
  Researcher: then how can one know?
  S10: frankly I don't know
  Researcher: what would you say?
  S10: I would say /frip/
  Researcher: why?

S10: I don't know I just feel the I here is not pronounced like you say it fast not /fraip/ you say it like you say <finish>

- *S13: <shrattec> /ʃrætik/ (then reads it multiple times to geminate the /t/ sound) … double T… /ʃrættik/. This C and H you say them sha and then the double T… it will make the word a bit difficult because you press on the T when you pronounce it and then you add a si at the end… but I would say /ʃrætik/. I expect a /k/… the pronunciation is that of /k/ not /s/.*
  Researcher: what made you expect that?
  S13: when I came to pronounce /ʃrætis/ I felt no it can't be.
  Researcher: Have you got a reason?
  S13: It's just my feeling, because the pronunciation of the word itself would be weird… /ʃrætis/…

However, on other occasions, that intuition was unsuccessful, as shown in the examples below:

- *S12: <knap> /knæp/ or maybe /næp/ because the K maybe silent because I came across a word but I don't remember what that word was and it was like that. The K was silent. So it could be /knæp/ or /næp/.*
  Researcher: which would you say is correct?
  S12: I don't know… I feel /nap/… but I also feel /knap/.

- *S12: <intuition> I know the T I O N is /ʃn/ it's pronounced like <action> so I think it's pronounced /inteʃn/*
  Researcher: and how about this part <u>?
  S12: we pronounce this like /inteʃn/ /intju:/… /intai/… I think /intaiʃn because there are words you feel that's how they should come. You can't think because you don't know much, so you just say /inteʃn/. I think so. Because I heard many words pronounced like that, so I think that's how they would be written. It's just me guessing.
  Researcher: what do you mean I can't think I guess?
  S12: but the word when it's new to you. It's not that you don't want to think but you have no idea so you expect you guess for example you heard it or you heard a word that is similar to it the whole thing will be guessing. I guessed. It's all guessing. Because I don't know the word. It's new to me. And when a word is new to me, I just guess. It could be right. It could be wrong. I also try to make it [its sound] assimilate words I already know.

- *S14: <hedfert> /hedfært/? I felt the E at the beginning is read as it is and the one at the end is read as /æ/.*

- *S14 <shrattec> /ʃʌtk/ the c is pronounced k here because there is no I or E or Y after it.*
  Researcher: and the beginning?
  S14: the C with the H that's sha
  Researcher: and how about the E before the C?
  S14: I didn't pronounce it I felt it would suit the word more if I leave it.

4.4.1.3.Other: Partial (incomplete) processing

A smaller number of responses (N=5, 1.3% of total responses) were coded in an "other" category. These responses appeared to reflect learners' engagement in some level of partial/incomplete processing of the target item. However, it should be noted that responses in this category often emerged in response to follow up questions/prompts from the researcher, as some of the examples below show:

- *S1: <Garment> "/græmint/ /gram/ and /mint/*
  Researcher: but I'm seeing /gær/ here
  S1: (attempts to read it again) /garment/ ah. No not /græmint/ yeah there is a difference in the letter position. I don't know I just went fast. The letter is there but I don't know I did not pay much attention.

- *S8: <hedfert> "/hedfret/ that's an easy word once I spell every letter I'd be able to say it. Its easy to pronounce too. /hedfret/. Simple letters.*
  Researcher: can you see it again
  S8: /hedfert/
  Researcher: and how come it was /hedfret/ first?
  S8: because I read it fast I felt it's easy but then when you said focus I noticed its ER.

- *S9: <Intuition> "intenʃin/ normal and this ending you say it /ʃin/ because there is I before it and after it.*
  Researcher: How about you look at it a second time
  S9: /intu:jiʃin/ I don't know, it's either intention or intuition
  Researcher: and how could one know
  S9: ask uncle google
  Researcher: why not you look at it
  S9: ah intuition because of the U
  Researcher: and why not <intention>
  S9: intention has not got a U
  Researcher: and how come it was intention first?
  S9: I don't know but intention has not got a U (and attempts to check the word and read it again)
  Researcher: and also intention has an N right
  S9: yes
  Researcher: and this one has not got that N?
  S9: I don't know. Did I read it intention? Maybe I did not concentrate … though I read it.

Finally, before concluding the findings into learners decoding processes, two points are worth noting. First, although the task was a retrospective-think aloud, on occasion, students engaged in concurrent think-aloud and the researcher allowed the students to proceed with their

concurrent verbalisation in those instances. Second, a handful of the responses were not given any codes as they were unclear and it was difficult to draw any interpretations from them.

4.4.2. Conclusion

In conclusion, analysis of learners' think-aloud protocols suggests that learners engaged in a range of processes when decoding L2 words. These include automatic processing, assembled processing (evident in small-unit processing, large-unit processing, and conscious computation of L2 GPCs by referencing to abstract L2 GPC knowledge, analogy-based processing, and/or intuition-guided processing) and also partial/incomplete processing. These processes did not operate exclusively from one another, as on many occasions they cooperatively contributed to the production of the final phonological form. In other words, learners tended to engage in more than one processing mechanism when decoding a single word/pseudoword.

At a micro level, each of these processes was associated with varying levels of success. In-depth analyses of the comments produced in each processing category provided further insights into the decoding difficulties experienced by these learners, such as: the presence of pre-existing inaccurate phonological forms; employing a small-unit processing mechanism when larger unit processing is required; difficulties associated with English syllabification rules; lack of L2 GPC knowledge; making inadequate analogies; relying on intuition rather than reason when converting graphemes to phonemes; and incomplete processing of the graphemes contained within a word.

Finally, it might be worth noting that there was a high degree of consistency in the processes reported by all 15 participants, in that each processing category (except the "other" category and the "intuition-guided processing" category) was observed across at least 80% of the sample. Further, no evidence was found for a primarily consonantal processing behaviour in the 360 comments analysed.

5. Results II: The contribution of second language decoding proficiency to second langauge reading comprehension

This chapter addresses the third research question: "What is the contribution of second language decoding proficiency to second language reading comprehension?" and is divided into two parts. The first part (section 5.1) deals with initial quantitative data screening and some preliminary analyses. In the second part (section 5.2), the research question is answered in depth. All statistical analyses reported in this chapter were conducted using SPSS (IBM Corp, 2015), R (R Core Team, 2018) and Microsoft Excel.

5.1. Data screening and preliminary analyses

5.1.1. Data screening:

5.1.1.1. Accuracy of input

The first step of the data screening process aimed at ensuring the accuracy of the data file. Accuracy of input of all variables was judged by plausible maximum, minimum, and mean values. A small number of incorrectly entered values were found and corrected.

5.1.1.2.Checking for missing data

Having checked the accuracy of the data file, the file was then screened for any missing data. Table 21 below presents the number of missing data for each of the variables.

Table 21. Amount of missing data on each of the continuous variables

variable	Cases on each measure (n)	Missing data (%)
L2 reading comprehension	233	0
L2 vocabulary knowledge	233	0
L2 grammar knowledge	227	6 (2.6%)
L1 reading comprehension ability	227	6 (2.6%)
L2 real word decoding proficiency	216	11 (4.8%)*
L2 pseudoword decoding proficiency	216	11 (4.8%)*
* Note: The figure is 11 (4.8%) and not 17 because only the cases who completed the L2 grammar knowledge and L1 reading comprehension tests (n= 227) were invited to complete the real word and pseudoword decoding tests. In other words, the 6 cases who did not attend session 2 were not invited to session 3.		

To determine how to deal with the missing data, the next step was to quantify their amount (in percentages) and observe their pattern. As the table above indicates, around 2.6% of the participants (n=6) who completed session 1 (i.e., the L2 reading comprehension and L2 vocabulary measures) had no data for session 2 (L2 Grammar knowledge and L1 reading comprehension ability). Furthermore, 4.8% of the participants (n=11) who completed session

2 had no data for session 3 (i.e., the L2 real word and L2 pseudoword measures). This pattern of loss was not unexpected, given that the data was collected across three sessions. Furthermore, the two-fold increase in the percentage of missing data in session 3 (i.e., from 2.6% in session 2 to 4.8%) was not unexpected. Unlike sessions 1 and 2, session 3 was not scheduled during the students' English class, and students often met with the researcher during their break hours to complete the session.

Various cut-off points regarding the acceptable proportion of missing data exist in the literature (for example, 5% in Schafer (1999), 10% in Bennett (2001)). Enders (2003) noted that a proportion of 15%- 20% is not uncommon in educational and psychological research. Given that only 7.4% (2.6% at time 2 and 4.8% at time 3) of the cases were missing, a decision was made to exclude these cases from further analyses.

5.1.1.3. Checking for univariate outliers

After listwise deletions of the cases with missing values, the dataset was then checked for univariate outliers. A univariate outlier is a case with an extreme score on one or more of the continuous variables. In order to detect these cases, z scores were calculated for each variable. Where a case had a z score that lay outside the ± 3.29 range, it was regarded as a potential outlier (Tabachnik and Fidell, 2014).

Using the cut-off value of ± 3.29, two cases were identified as univariate outliers. The first was a case that had a z score of 5.50 on the second language vocabulary measure. The second univariate outlier was a case that had a z score of -5.45 on the L1 reading comprehension measure. In addition to the inspection of z scores, histograms were also examined for each variable. In line with the z score findings, the two cases above were the only two observations that were disconnected from the rest of the distribution.

Once these outliers were identified, a decision had to be made as to how to deal with them, i.e., by deleting, rescoring, or variable transformation. According to the ESL institute, Case 1 was classified as an A1 level student. Interestingly, however, this case had the highest scores on all the variables. Her scores were not only higher than all A1 level students, but also higher than all B1 level students. Thus, clearly this case does not represent typical foundation year ESL students. The researcher's previous work experience in this context allowed her to speculate as to why this unusual observation was present. Though not the usual practice, a foundation year student might choose not to apply for exemption from the ESL course and 'intentionally' underperform on the placement test. This intentional underperformance is often driven by the student's desire to be admitted in her first major of choice at the end of the foundation year. Registering for ESL courses would allow the highly proficient ESL student to

obtain excellent grades on the ESL course. This in turn would increase her overall GPA and eventually increase her chances of admittance into her first major of choice. Another possible reason would be students' interest in English. Registering for the ESL course would provide students interested in English with opportunities to pursue their interest further. Which of these two explanations, if any, is true is not important. What is important is that this case clearly does not represent the population of low proficiency ESL students. A decision was therefore made to remove the case from the dataset.

As for case 2, it is unclear why this observation was far from the rest of the distribution. However, it is worth noting that the recorded score for this case was zero. This was considered very unusual given that this score was achieved on the Arabic reading comprehension measure. As this score does not represent the Arabic comprehension ability of college level students, a decision was made to remove the case.

Following the deletion of these two cases, z scores were recalculated to check if other outlier cases were masked by the presence of the above two outliers. No further outliers were identified.

5.1.1.4.Checking for multivariate outliers

After checking for univariate outliers, the data was checked for multivariate outliers. A multivariate outlier is a case with an abnormal combination of scores on two or more variables (Tabachnick & Fidell, 2014). Multivariate outliers are problematic, as their presence could change the line of best fit, affecting regression estimates and causing large errors (Field, 2015). Therefore, it was important to identify multivariate outlier cases and deal with them.

Various methods exist for the detection of multivariate outliers. These include examining Mahalanobis distance, standardised residuals, leverage values, and Cook's distance. Of these methods, Mahalanobis distance remains the most widely used method for outlier detection (Tabachnick & Fidell, 2014).

Mahalanobis distance (MD) measures the influence of a case by examining its distance from the centroid (overall mean of the various predictor variables). A large MD value would indicate that the case is far from the centroid, and as such a potential multivariate outlier. To establish a cut-off value, researchers find the critical value of chi-square for their desired alpha level, for example, $p = .05, .01$ (Field, 2015), or .001 (Tabachnick & Fidell, 2014). It is worth noting, however, that the reliability of the MD measure has been heavily questioned (Egan & Morgan, 1998; Hadi & Simonoff, 1993; Leys, Klein, Dominicy, & Ley, 2018). In their Monte Carlo study, Leys, Klein, Dominicy, and Ley (2018) documented that the most accurate

detection level for MD is the chi-square at p=.01. Thus, in the present study, any case with a distance greater than 13.28 (p = .01) was considered a potential multivariate outlier.

Two cases (i.e., less than 1% of the total population) had a MD that exceeded this cut-off value. The first case, hereafter referred to as case 137, had a MD of 21.65 and the second, hereafter referred to as case 111, had a MD of 15.28.

When multivariate outliers are detected using MD, it is common practice to examine the cases further by looking at their Cook's distance, standardised/studentised residuals, and leverage values before any deletion decision is made. Therefore, these three metrics were checked for the above two cases.

Cook's distance measures the overall influence of a case on the model. To detect a problematic case, various numerical and graphical methods have been suggested in the literature. Numerical methods have produced various cut-off points at which to consider a case as problematic. A popular cut-off point is calculated using the formula 4/N, where N is the sample size. In the present study, this formula would yield a threshold of ~.02 (i.e., 4/214). When the Cook's distance of each case was obtained, case 111 appeared to have significantly exceeded this threshold with a distance score of ~.19. Case 137 did not, with a distance of ~.01.

Rather than relying on numerical thresholds, Fox (1991) and Chatterjee and Hadi (2015) recommend plotting a graph and looking for cases whose Cook's distance appear to be substantially larger than the rest of the cases. If a case with relatively high Cook distance is found, that case would be a potential outlier even if it does not exceed the numerical cut-off value. Figure 32 below again shows that case 111 appears largely disconnected from the rest of the cases.

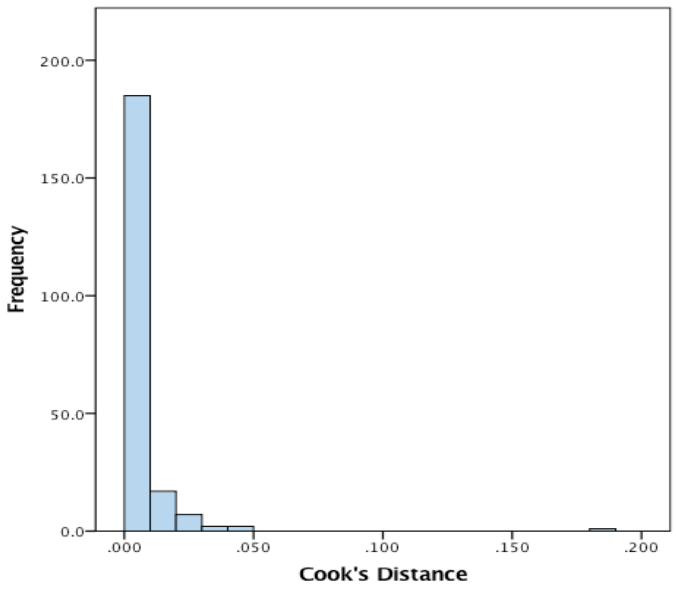


Figure 31. Cook's distance of the 214 cases

The leverage values for the two cases were also calculated. Leverage measures the influence of the observed value of the dependent variable over the predicted values. The average leverage value is (k+1)/n where k is the number of predictors in the model and n is the number of the cases. Leverage values can lie between 0 (the case has no influence) and 1 (the case has complete influence over prediction) (Field, 2015). If no cases exert undue influence over the model, then all leverage values are expected to be close to the average value (ibid).

Stevens (2002) recommends using three times the average (3(k+1)/n) as a cut-off point for identifying cases which exert undue influence over prediction. In the present study, this formula would yield a threshold of ~.07. When the leverage values of each case were calculated, cases 137 and 111 appeared to be the only two problematic cases, with the former obtaining a value of ~.10 and the latter a value of ~.07.

Standardised and studentized residuals were also checked. These residuals represent the error present in the model. (For the purpose of outlier detection, studentised residuals are preferable, as they provide a more precise estimate of the error variance of a specific case (Field, 2015)). Cases with residuals greater or smaller than ±3.29 would be a cause of concern. In the present study, only one case appeared to exceed this threshold, case 111, with a studentised residual of -3.44 and a standardised residual of -3.31. Case 137 did not appear to be problematic with a standardised residual of -0.42 and a studentised residual of -0.44.

In short, according to Mahalanobis distance, only two of the 214 cases (i.e., less than 1%) appeared to be multivariate outliers. Furthermore, using various techniques of multivariate outlier analysis, case 111 emerged as an outlier in each subsequent analysis (Cook's distance, leverage analysis, standardised and studentised residuals). Case 137 emerged as an outlier in two analyses (leverage analysis and Mahalanobis distance) but did not appear to be problematic in the other analyses (see table 22 below). In light of these findings, and because these two cases constitute less than 1% of the sample, a conservative decision was made to delete both cases.

Table 22. Mahalanobis distance, Cook's distance, standardised residuals, Studentised residuals, and Leverage values for cases 111 and 137

Case	Mahalanobis distance (13.28)	Cook distance (.02)	Standardised residual (±3.29)	Studentized residual (±3.29)	Leverage value (.07)
111	15.20	.19	-3.31	-3.44	.07
137	21.65	.01	-0.42	-0.44.	.10

*Notes: *cut-off values are shown in parenthesis. **values exceeding cut-off values are underlined.*

5.1.1.5. Checking for normality

Although normality of variables' distributions is not an assumption in regression analyses (Tabachnick & Fidell, 2014), establishing normality was considered a useful preliminary step. This is because normally distributed variables could enhance the predictability of multivariate regression models (ibid). Therefore, all variables measured in the present study were screened for normality. (Assumptions specific to the multivariate analytical techniques used to answer the research questions are addressed in part 2 of this chapter)

Various methods exist for assessing normality. These include running normality tests, visual inspection of histograms, and examination of skewness and kurtosis z scores. In large samples, traditional normality tests (i.e., Kolmogorov-Smirnov and Shapiro-Wilk) have been documented to produce significant results even when distributions appear normal or exhibit minor deviations from normality (Field, 2015; Kim, 2013; Tabachnick & Fidell, 2014). Therefore, in the present study, following an initial inspection of histograms, normality was checked using skewness and kurtosis z values.

Initial screening of histograms suggested that all variables were normally distributed except two: the L2 vocabulary knowledge variable, and possibly, the L2 reading comprehension variable. Nonetheless, a decision was made to further assess the distribution of all variables statistically using skewness and kurtosis z values.

Z scores were obtained by dividing the skew and kurtosis values of each variable by their standard errors. Because standard errors tend to be small in large sample sizes (leading to significant results using formal normality tests (Field, 2015; Kim, 2013)), Kim (2013) recommends using different cut-off values for different sample sizes. For sample sizes ranging between 50 and 300, z scores which lay out of the ±3.29 range would be indicative of non-normality (at $p < .05$) (ibid). Table 23 below shows the skewness and kurtosis z scores of all the variables measured in the present study.

Table 23. Skewness and kurtosis z scores of all the study variables.

	N	Skewness	Std. Error	Kurtosis	Std. Error	skewness z score	kurtosis z score
L2RC	212	-0.66	0.17	-0.10	0.33	**-3.98**	-0.30
L2VK	212	0.97	0.17	0.87	0.33	**5.80**	2.60
L2GK	212	-0.15	0.17	-0.81	0.33	-0.90	-2.43
L1RC	212	-0.27	0.17	-0.21	0.33	-1.62	-0.62
L2DP	212	-0.08	0.17	-0.68	0.33	-0.46	-2.05
L2RDP	212	-0.31	0.17	-0.74	0.33	-1.85	-2.23
L2PDP	212	0.30	0.17	-0.56	0.33	1.78	-1.67

Notes:
*L2RC (L2 reading comprehension), L2VK (L2 vocabulary knowledge), L2GK (L2 grammar knowledge), L1RC (L1 reading comprehension ability), L2DP (L2 decoding proficiency), L2RDP (L2 real word decoding proficiency), L2PDP (L2 pseudoword decoding proficiency)
**figures in bold are suggestive of departures from normality (at $p < .05$)

In line with the eyeball test, the skewness and kurtosis *z*-values of the L2 grammar knowledge, L1 reading comprehension ability, L2 decoding proficiency, L2 pseudoword decoding proficiency, and L2 real word decoding proficiency were all well within the acceptable ±3.29 range, suggesting normality of distribution. The skewness *z*-scores of the L2 reading comprehension and the L2 vocabulary knowledge measures, however, have exceeded that range, with *z*-values of -3.98 and 5.80, respectively. Because the L2 reading comprehension measure deviated only slightly from normality, a decision was made to not transform the variable. However, deviation from normality was moderate in the L2 vocabulary knowledge measure, indicating a need to transform the variable.

The next step was to decide on which transformation method to apply. Field (2015) recommends trialing different transformation methods until the most helpful transformation is found, i.e., the one that yields the skewness and kurtosis values nearest zero. Tabachnick and Fidell (2014) recommend applying square root transformations when deviation from normality is moderate. When it is severe, a logarithmic transformation is recommended (ibid). A visual inspection of the histograms of the L2 vocabulary knowledge variable suggested that the deviation was only moderate, suggesting the suitability of the square root method. However, following Field's recommendation, a decision was made to apply both transformation methods (i.e., square root and logarithmic) and examine the results they generate. Because the smallest score on the L2 vocabulary measure was zero, a constant (1) was added to each score prior transformation so that that smallest score is 1.

Skewness and kurtosis z values were then calculated for the vocabulary knowledge variable. As can be seen from table 24 below, the square root method appears to have produced the most helpful results, as z values were well within the recommended range. Clearly, because the distribution was only moderately non-normal, the logarithmic transformation was too powerful, so that the distribution became negatively skewed (when it was initially positively skewed). Therefore, the square-root transformed variable was chosen to be used in subsequent analyses.

Table 24. Skewness and kurtosis values and z scores prior transformation of the L2 vocabulary knowledge measures, post square root transformation, and post logarithm transformation

pre transformation							
Measure	N	skewness			kurtosis		
		statistic	std. error	*z*-score	statistic	std. error	*z*-score
L2 vocabulary knowledge	212	0.969	0.167	**5.80**	0.865	0.333	2.60
post transformation (Square root)							
Measure	N	skewness			kurtosis		
		statistic	std. error	*z*-score	statistic	std. error	*z*-score
L2 vocabulary knowledge	212	0.186	0.167	**1.11**	-0.236	0.333	-0.71
post transformation (logarithm)							
Measure	N	skewness			kurtosis		
		statistic	std. error	*z*-score	statistic	std. error	*z*-score
L2 vocabulary knowledge	212	-0.788	0.167	**-4.72**	0.854	0.333	2.56

5.1.2. Descriptive statistics

5.1.2.1. Descriptive statistics for the background questionnaire

As described in section 3.4.1.1.6, though the background questionnaire data played no formal role in answering the research questions, it served two purposes. Firstly, the data provided a basis to verify the eligibility of the participants. Secondly, the data provided information on the other languages that students were familiar with, their proficiency level in those languages, and whether they had received any reading instruction in those languages. As

expected, no participant was judged to be ineligible to participate[20] (see section 3.4.1.1.6). Data from questions that yielded variable responses are presented in table 25 below:

Table 25. Descriptive statistics for the background questionnaire

Age	Average: 19 Min: 18 Max: 28		
Do you know any language other than Arabic and English?	No (77.8%)		
	Yes (22.2%)	How many languages do you know other than English and Arabic?	One (78.7%) Two (19.1%) Three (2.1%)
		What are those languages?	Turkish (65.5%) Korean (18.2%) French (7.3%) Spanish (3.6%) Hindi (3.6%) German (1.8%)
		What is your proficiency level in that language?	Beginner (97.9%) Intermediate (2.1%)

As the table above shows, 22.2% of the participants indicated knowledge of a language other than English and Arabic. When this was the case, this language was Turkish (65.5%), Korean (18.2%), French (7.3%), Spanish (3.6 %), Hindi (3.6%), or German (1.8%). Proficiency level in those languages was described as beginner (97.9% of the time), and intermediate (2.1% of the time), and never advanced (0%). It might be worth mentioning here that the researcher often engaged in brief discussions with these students to know more about their language backgrounds. No participant indicated having formally learnt the third language, as almost all noted that they had "picked up some vocabulary/phrases" from watching movies/series in that language. That was not unexpected, as English is the only foreign language taught in Saudi state schools.

Further, the researcher asked the one case who indicated that her knowledge of the third language was intermediate: "Would you say you are more proficient in Turkish than English?", with a view to check if her self-rating was reliable. The participant noted that she is much more proficient in English than Turkish. Interestingly, this student was classified as an

[20] Two students mentioned that they attempted to learn how to read the Turkish alphabet from YouTube. However, when they were asked basic questions about Turkish decoding, one was unable to provide any correct answer. The other made one successful attempt. Nonetheless, the researcher judged her knowledge to be extremely basic. Thus, neither of these two participants was judged to be able read in the third language, and both cases were retained.

A2/elementary student. This suggests that this student might not have been able to self-rate her proficiency in the third language accurately.

All in all, these findings verified the eligibility of the present study participants. Further, the proficiency level in the third language was almost always beginner. This suggests a likely minimal, if any, influence of that language on learners' decoding realisations in English. Nonetheless, it remains a possibility that the decoding realisations of these participants could be influenced by the phonemic inventory of a language other than Modern Standard Arabic and Saudi Arabic. To ensure the researcher's phonological transcriptions are not influenced by her lack of familiarity to certain sounds, the researcher sought to familiarise herself with the phonemic inventories of the languages she is not familiar with.

5.1.2.2. Descriptive statistics for L2 reading comprehension, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, and overall L2 decoding proficiency measures

Table 26 below presents the descriptive statistics for all variables.

Table 26. Descriptive statistics for the study variables

measure	N	mean	SD	max	min
L2 reading comprehension	212	16.55	4.68	24	3
L2 vocabulary knowledge	212	11.24	7.41	36	1
L2 grammar knowledge	212	17.17	6.21	30	4
L1 reading comprehension ability	212	14.01	2.34	19	8
L2 real word decoding proficiency	212	49.59	10.89	71	24
L2 pseudoword decoding proficiency	212	24.82	9.18	48	7
Overall L2 decoding proficiency (a composite measure of L2 real word and L2 pseudoword decoding proficiency)	212	74.42	18.9	117	33

5.2. Answering the research question: What is the contribution of second language decoding proficiency to second language reading comprehension performance?

In order to answer this question in depth, the following seven issues were investigated in turn:

1) The relationships between overall L2 decoding proficiency, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension ability (section 5.2.1);
2) The contribution of overall L2 decoding proficiency to L2 reading comprehension *before* controlling for the contributions of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability (section 5.2.2);
3) The contribution of overall L2 decoding proficiency to L2 reading comprehension *after* controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability (section 5.2.2);
4) The contribution of overall L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the L2 learner (higher level of linguistic proficiency vs. moderate level of linguistic proficiency vs. lower level of linguistic proficiency) (section 5.2.3);
5) The independent contributions of L2 real word decoding proficiency and L2 pseudoword decoding proficiency to L2 reading comprehension (section 5.2.4);
6) The relative importance of L2 decoding skill for L2 reading comprehension: insights from Dominance Analysis and Relative Weight Analysis (section 5.2.5), and finally,
7) The contribution of L2 decoding proficiency to L2 reading comprehension: qualitative insights from the experiences of second language readers (section 5.2.6).

The rationale for investigating each of these issues was discussed in section 3.5.2.

5.2.1. The relationships between L2 reading comprehension performance, L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 decoding proficiency, L2 real word decoding proficiency, and L2 pseudoword decoding proficiency

Prior to running the correlational analysis, the data was checked for linearity graphically using scatterplots of all the variables. The plots showed that there was a linear relationship among all correlated variables, suggesting the suitability of a Pearson's correlation test (Normality was screened for and discussed in section 5.1.1.5). However, as the L2 reading comprehension measure showed a very slight deviation from normality, Spearman correlations (table 27) were conducted (though note that Pearson and Spearman correlations produced nearly equivalent coefficients. Pearson correlations are found in table 43 (Appendix L) for further reference). The results from Spearman correlations are discussed below.

Table 27. Spearman correlations between study variables.

	L2 reading comprehension	L2vocabulary knowledge	L2 grammar knowledge	L1 reading comprehension ability	L2 decoding proficiency	L2 real word decoding proficiency	L2 pseudoword decoding proficiency
L2 reading comprehension	1	.74**	.74**	.14*	.59**	.63**	.48**
L2vocabulary knowledge (transformed)		1	.77**	.04	.66**	.73**	.49**
L2 grammar knowledge			1	.06	.62**	.70**	.43**
L1 reading comprehension ability				1	.02	-.01	.05
Overall L2 decoding proficiency					1	***	***
L2 real word decoding proficiency						1	.77**
L2 pseudoword decoding proficiency							1

Notes:
* Correlation is significant at the .05 level (2-tailed).
** Correlation is significant at the .01 level (2-tailed).
*** Overall L2 ddecoding proficiency is a composite measure of real word and pseudoword decoding proficiency. Thus, the correlation between L2 decoding proficiency and L2 real word decoding proficiency will intuitively be inflated as the former is partly constructed by readers' scores on the latter. The same applies for the correlation between L2 decoding proficiency and L2 pseudoword decoding proficiency. For this reason, these correlations are omitted from the table.

As can be seen from table 27 above, L2 reading comprehension performance was significantly related to L2 vocabulary knowledge, $r = .74$, $p < .001$, L2 grammar knowledge, $r = .74$, $p < .001$, L1 reading comprehension ability, $r = .14$, $p = .046$, overall L2 decoding proficiency, $r = .59$, $p < .001$, L2 real word decoding proficiency, $r = .63$, $p < .001$, and L2 pseudoword decoding proficiency, $r = .48$, $p < .001$. Noticeably, almost all of these coefficients are medium to large in magnitude, except that between L1 reading comprehension ability and L2 reading comprehension which, though (borderline) statistically significant, appears quite small.

Though correlations involving the reading comprehension variable are prioritised in the present work, some correlations are also noteworthy. These include those between L2 grammar knowledge and L2 vocabulary knowledge, $r = .76$, $p < .001$, L2 vocabulary knowledge and L2 decoding proficiency, $r = .65$, $p < .001$, L2 vocabulary knowledge and L2 real word decoding proficiency, $r = .73$, $p < .001$, and L2 pseudoword decoding proficiency and L2 real word decoding proficiency, $r = .77$, $p < .001$. The fact that these variables have shown strong relationships with reading comprehension suggest that they are likely to be significant predictors of L2 reading comprehension performance. However, the fact these variables have also shown strong relationships among themselves highlights the need for rigorous evaluation of their *unique* contributions in predicting reading comprehension performance, which the following sections aim to address in depth.

5.2.2. The contribution of overall L2 decoding proficiency to L2 reading comprehension *before* and *after* controlling for L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability

Although the stringent test for the contribution of L2 decoding proficiency – and the one the researcher supports – would involve controlling for other reading comprehension predictors, it was deemed useful to report at the outset the contribution of L2 decoding proficiency prior to controlling for these variables.

Therefore, a hierarchical regression analysis was initially implemented to examine the contribution of decoding to reading comprehension before taking into account the effects of other reading comprehension subskills. Overall L2 decoding proficiency (a composite measure of L2 real word and L2 pseudoword decoding), L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability were entered as predictor variables, and L2 reading comprehension as a dependent variable. The sequence of entry is shown below:

Table 28. Order of variable entry

Step 1	L2 decoding proficiency (a composite measure of L2 real word and L2 pseudoword decoding proficiency)
Step 2	L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability

Following an initial run of the model, the data was screened to see if it met the assumptions of multiple regression (normality, homoscedasticity, linearity, and independence of residuals, and absence of collinearity among the model predictors). To evaluate the assumption of normality of residuals, a histogram of standardised residuals was plotted. As can be seen in figure 33, the residuals appear normally distributed. Kolmogorov-Smirnov and Shapiro-Wilk tests were also used for the same purpose. The distribution of residuals in both tests was found to be normal, $p > .05$.

To evaluate the assumption of homoscedasticity of residuals, a scatterplot of standardized residuals against predicted DV values was produced. As can be seen in figure 35, the residuals' variance appears equal. A Breusch-Pagan test was also run for the same purpose. The variance of residuals was found to be equal, $p > .05$.

Visual inspection of a p-p plot of standardized regression residuals suggest that the linearity assumption has been met (Figure 34). To evaluate the independence of errors

assumption, the Durbin-Watson test was used. The test statistic was 2.019, suggesting that the errors are independent (Field, 2015).

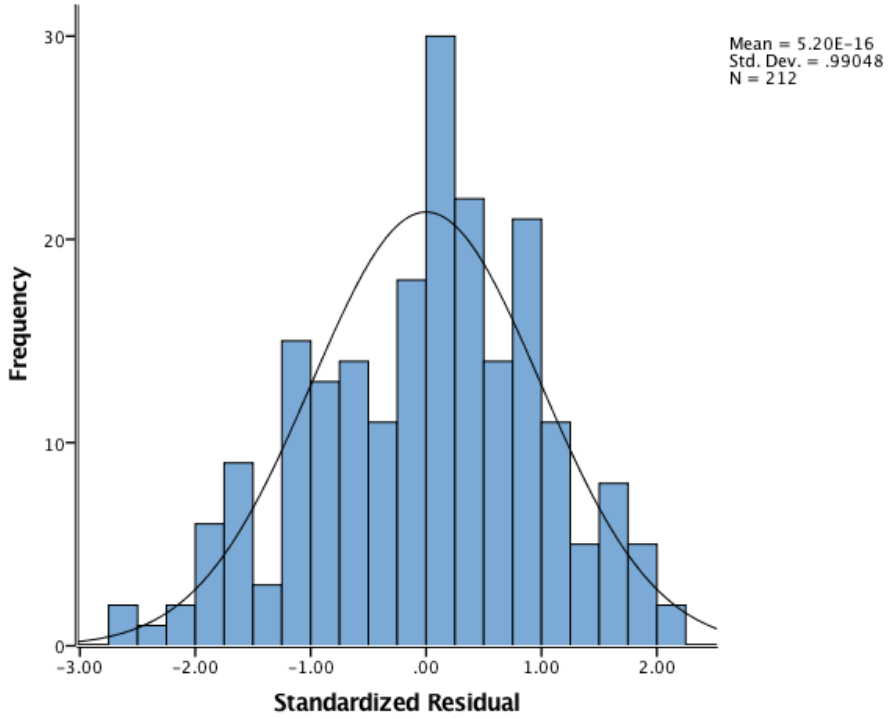


Figure 32. Histogram of normality of residuals

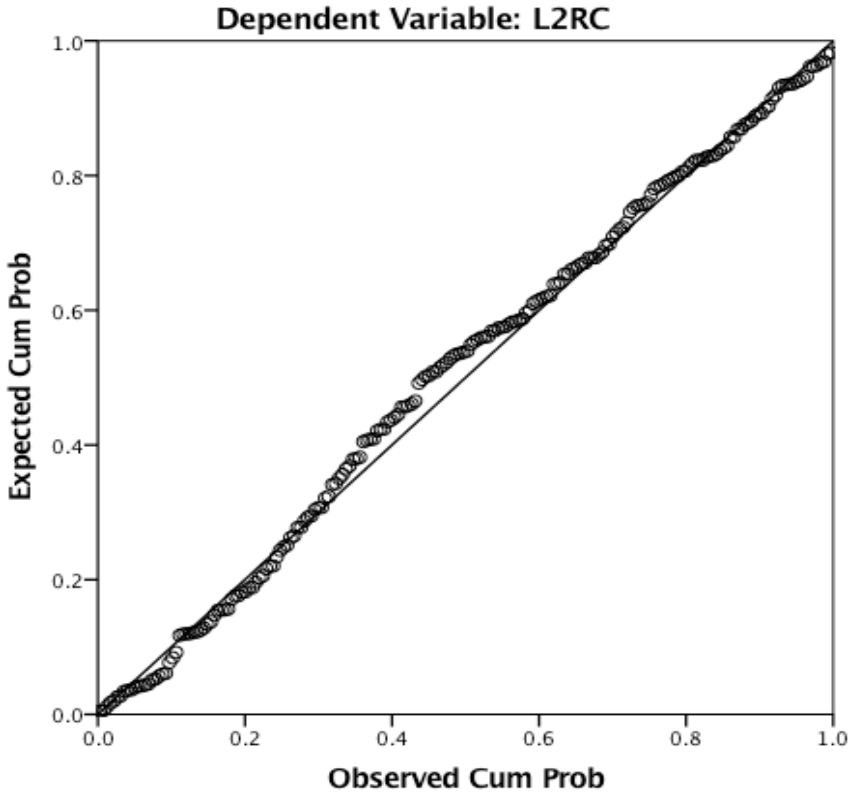


Figure 33. Normal p-p plot of regression standardised residuals

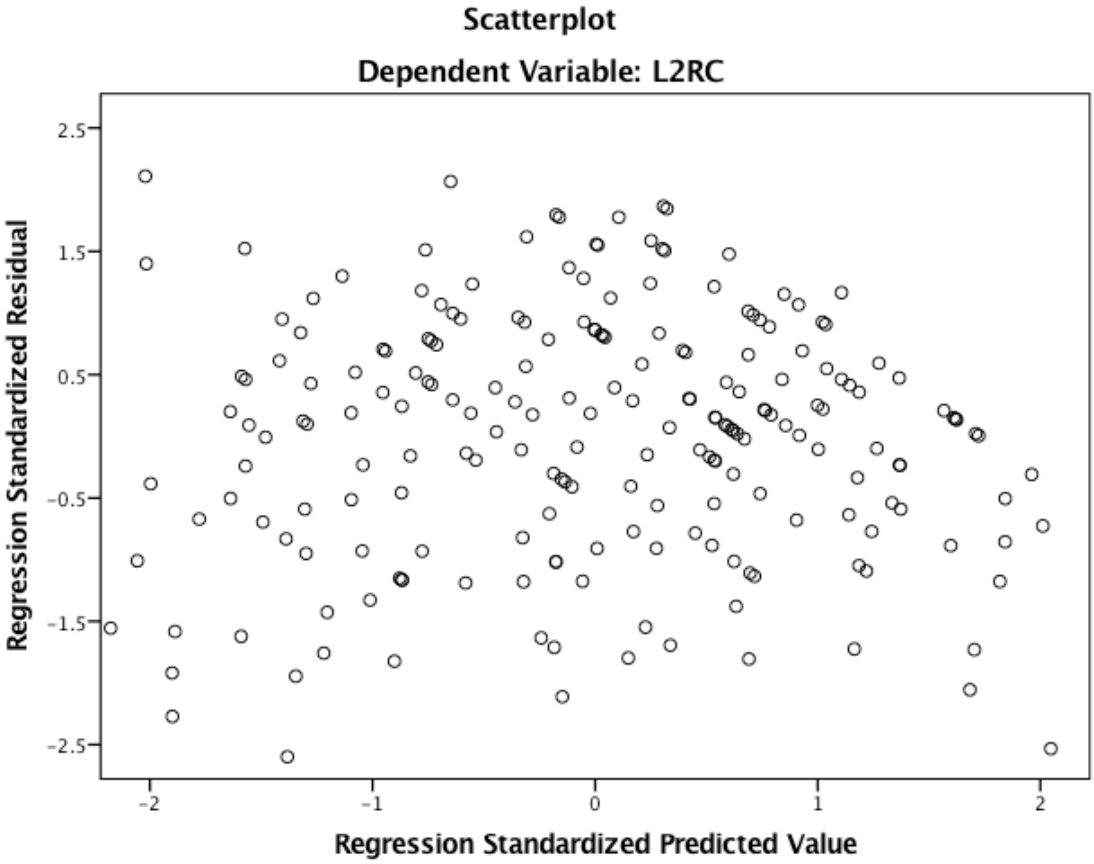


Figure 34. Scatterplot of residuals against predicted reading comprehension scores

Finally, an examination of correlations (see Table 27) suggests that no independent variables were seriously correlated, in that no correlation reached the threshold of .9 (ibid). However, because L2 grammar knowledge and L2 vocabulary knowledge (and possibly L2 decoding proficiency and L2 vocabulary knowledge) were somewhat highly correlated, the assumption of multicollinearity was assessed using additional collinearity diagnostics. Again, all figures were well within acceptable limits. That is (a) none of the variables had a variance inflation factor (VIF) of 10 or larger (the largest VIF value was 2.680) or a tolerance of .2 or less (the smallest tolerance value was .373); (c) no conditioning index was greater than 30 (the largest conditioning index was 19.97) and (d) variance proportions were not greater than .50 for any two variables.

Now that all the assumptions of multiple regression have been met, the statistics of the generated model are reported to address the first issue of interest in this section: *What is the contribution of L2 decoding proficiency to L2 reading comprehension before taking into account the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability?* Because this section is mainly concerned with the contribution of decoding to reading comprehension independently of the effects of other predictors, the

discussion below focuses on the results of the first model (step1) and mainly the sequential R^2 of the second (final) model.

From a hierarchical multiple regression analysis with reading comprehension as the criterion variable, and overall L2 decoding proficiency as a predictor entered at stage 1, a statistically significant model emerged (Adjusted R^2= .357, F (1, 210)=117.900, p<0.001). The adjusted R^2 value indicate that more than a third (35%) of the variability in L2 reading comprehension performance is predicted by decoding proficiency. Introducing the (square root) of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability explained an additional 28% of variation in reading comprehension and this change in R^2 was significant, F (4, 207) = 91.079, p < .001. Together, the four independent variables accounted for 63% of the variance in L2 reading comprehension performance (Table 46, appendix L). (To avoid redundancy, the standardized regression coefficients of all variables – in the final model– are presented in table 29 below, along with the sequential R (R change) value of L2 decoding proficiency when entered at the final step).

Now, it might be tempting to conclude that L2 decoding proficiency explained a third of the variance in reading comprehension. However, as mentioned previously, this approach is problematic for several reasons. Therefore, a second hierarchical regression analysis was performed to see if prediction of L2 reading comprehension is enhanced by the inclusion of decoding proficiency *after* adjustment for variations in L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability.

As can be seen from Table 29 below, at stage one, L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability contributed significantly to the regression model, F (3,208) = 117.807 p< .001, and accounted for 63% (~62% adjusted) of the variability in L2 reading comprehension ability. Adding L2 decoding proficiency to the regression model explained an additional ~1% of the variation in reading comprehension and this change in *R2* was significant, F (4, 207) = 91.079, p < .05. In this final model, the four variables explained ~64% (~63% adjusted) of the variability in L2 reading comprehension ability.

One method to evaluate the relative importance of a variable, i.e., the size of its contribution relative to other variables in the model, is to examine standardised beta values (Field, 2015). As shown in table 29 below, L2 grammar Knowledge (ß=.428, p<.001) emerged as the strongest predictor of L2 reading comprehension, followed by L2 vocabulary knowledge (ß=.314, p<.001). Overall L2 decoding emerged as the third strongest predictor of L2 reading

comprehension (ß=.124, p<.05), followed by L1 reading comprehension ability (ß=.108, p<.05).

Table 29. Summary of hierarchical regression analysis for variables predicting L2 reading comprehension

	B	**ß**	**T**	**R**	**R^2**	**Adjusted R^2**	**ΔR**
Step 1				.793	.630	.624	.630***
L2 vocabulary knowledge	1.641	.371	1.545***				
L2 grammar knowledge	.347	.461	5.832***				
L1 reading comprehension ability	.219	.109	7.215*				
Step 2				.799	.638	.631	.008*
L2 vocabulary knowledge	1.387	.314	4.579***				
L2 grammar knowledge	.322	.428	6.571***				
L1 reading comprehension ability	.216	.108	2.571*				
Overall L2 decoding proficiency	.031	.124	2.160*				
Note. N = 212; *p < .05, **p < .01, ***p< .001							

In conclusion, L2 decoding proficiency made a significant contribution to L2 reading comprehension both before *and* after the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability were controlled for. This contribution appeared to be larger in magnitude than that of L1 reading comprehension ability, yet smaller than the contributions of L2 grammar knowledge and L2 vocabulary knowledge.

5.2.3. The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner

It was of interest to see whether the contribution of L2 decoding proficiency varies in readers at differing levels of L2 linguistic proficiency (As mentioned previously, linguistic proficiency was represented using a combined measure of L2 vocabulary and L2 grammar knowledge). Therefore, a moderated multiple regression was conducted using the PROCESS macro for SPSS developed by Hayes (2013; 2018). The model included L1 reading comprehension ability, overall L2 decoding proficiency, L2 linguistic proficiency (the moderator variable), and the interaction term L2 decoding proficiency*linguistic proficiency (As mentioned previously (section 3.5.2.3), it would be meaningless to include L2 vocabulary knowledge and L2 grammar knowledge, since the linguistic proficiency variable, which is a composite measure of both, is already entered in the model.

As can be seen in table 30, there was a significant proficiency*decoding interaction, b=–.005, t=–3.151, p <.01. This result suggests that the contribution of L2 decoding proficiency to L2 reading comprehension decreased as learners' linguistic proficiency increased.

Table 30. Summary of hierarchical moderated regression analysis

	B	**ß**	**T**	**sr²**	**R**	**R²**	**Adjusted R²**	**ΔR**
Step 1					.78	.60	.60	.60***
L1 reading comprehension ability	.203	.102	2.319*	.101				
L2 linguistic proficiency (centered)	.505	.760	17.314***	.757				
Step 2					.79	.62	.61	.02**
L1 reading comprehension ability	.203	.102	2.371*	.101				
L2 linguistic proficiency (centered)	.428	.643	11.526***	.493				
L2 decoding proficiency (centered)	.045	.183	3.280**	.140				
Step 3					.80	.64	.63	.02**
L1 reading comprehension ability	.166	.083	1.957	.082				
L2 linguistic proficiency (centered)	.433	.650	11.899***	.498				
L2 decoding proficiency (centered)	.039	.158	2.865**	.120				
L2 linguistic proficiency (centered) * L2 decoding proficiency (centered)	-.005	-.135	-3.151**	-.132				
Note. $N = 212$; *$p < .05$, **$p < .01$, ***$p < .001$								

To visualize the moderation effect, the interaction was probed at high (+1 SD above the mean), average (mean) and low (–1 SD below the mean) levels of the moderator (Table 31 and Figure 36). This is the most popular approach to probing interactions in regression models involving a continuous moderator (Aiken & West, 1991; Hayes, 2013), and is the method used

in PROCESS. When proficiency level was low, decoding proficiency was a significant predictor of reading comprehension, with b= .074, t= 4.544, $p < .001$. When proficiency level was moderate, decoding proficiency was also a significant predictor of reading comprehension, though the size of the contribution was lower, with b= .039, t= 2.865, $p < .01$. When proficiency level was higher, the contribution of decoding to reading comprehension was not significant, with b= .004, t= .204, $p < .838$.

Table 31. Summary of follow up simple slope analysis

Linguistic proficiency	B	t	*p*
Low (-7.032)	.074	4.544	.000
Average (.000)	.039	2.865	.005
High (7.032)	.004	.204	.838

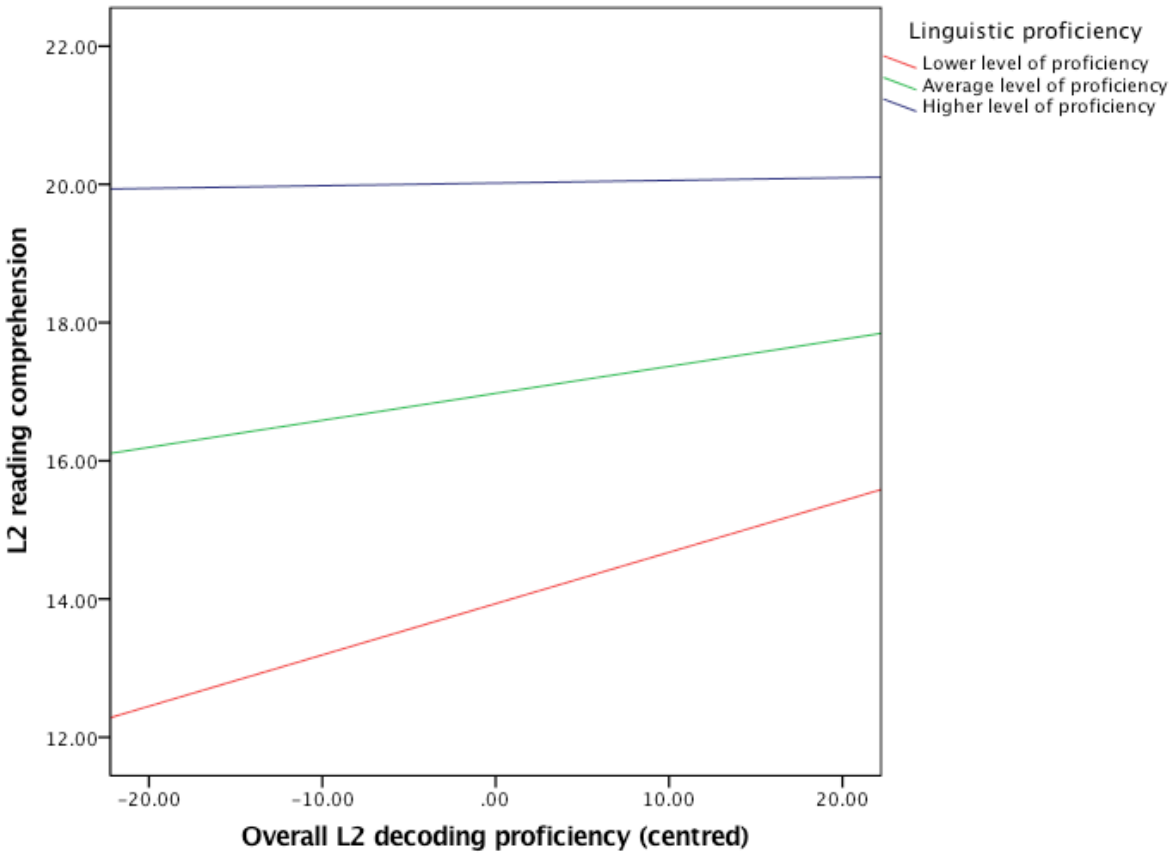


Figure 35. The contribution of overall L2 decoding proficiency as a function of the proficiency level of the learner

5.2.4. The contribution of second language decoding proficiency to second language reading comprehension as a function of the decoding measure

To address this issue, two sets of analyses (A and B) are reported below.

- **Analysis A:**

Table 32 (next page) presents an overall summary of the findings of the first set of analyses in this section (Analysis A), which aimed at: (1) examining the contribution of pseudoword decoding proficiency to L2 reading comprehension, before and after the contributions of L2 linguistic proficiency and L1 reading comprehension ability have been partialled out **(analyses A1 and A2 respectively)**; (2) examining whether pseudoword decoding proficiency predicts any variations in L2 reading comprehension after additional controlling for L2 real word decoding proficiency **(analysis A3)**; and (3) examining whether the contribution of L2 pseudoword decoding proficiency would vary as a function of the linguistic proficiency of the learner, and if yes, how **(analysis A4).**

Table 32. Summary of regression analyses for investigating the contribution of L2 pseudoword decoding proficiency to L2 reading comprehension

Analysis A							
	B	**ß**	**T**	**R**	**R^2**	**Adj R^2**	**ΔR**
Analysis A1							
Step 1				.47	.22	.21	.22***
L2 pseudoword decoding proficiency	.238	.467	7.651***				
Step 2				.79	.62	.61	.40***
L2 pseudoword decoding proficiency	.073	.143	2.963**				
L1 reading comprehension ability	.195	.097	2.259*				
L2 linguistic proficiency	.462	.695	14.371***				
Analysis A2 (order of entry switched)							
Step 1				.78	.60	.60	.60***
L1 reading comprehension ability	.203	.102	2.319*				
L2 linguistic proficiency	.505	.760	17.314***				
Step 2				.79	.62	.61	.016**
L1 reading comprehension ability	.195	.097	2.259*				
L2 linguistic proficiency	.462	.695	14.371***				
L2 pseudoword decoding proficiency	.073	.143	2.963**				
Analysis A3 (additional controlling for L2 real word proficiency)							
Step 1				.79	.62	.61	.62***
L1 reading comprehension ability	.213	.107	2.479*				
L2 linguistic proficiency	.411	.618	9.879***				
L2 real word decoding proficiency	.083	.194	3.110**				
Step 2				.79	.62	.61	.00
L1 reading comprehension ability	.205	.103	2.379*				
L2 linguistic proficiency	.422	.635	9.836***				
L2 real word decoding proficiency	.054	.126	1.397				
L2 pseudoword decoding proficiency	.037	.073	1.046				
Analysis A4 (introducing the interaction term)*							
Step 2				.79	.62	.62	.01
L1 reading comprehension ability	.177	.088	2.048*				
L2 linguistic proficiency	.461	.693	14.405***				
L2 pseudoword decoding proficiency	.074	.145	3.010**				
L2 PDP*linguistic proficiency	-.006	-.080	-1.857				
Note. $N = 212$; *$p < .05$, **$p < .01$, ***$p < .001$							

As can be seen in the table above (Analysis A1), from a hierarchical multiple regression analysis with reading comprehension as the criterion variable, and L2 pseudoword decoding proficiency as a predictor entered at stage 1, a statistically significant model emerged (Adjusted R^2= .21, F =, p<0.001), suggesting that L2 pseudoword decoding proficiency explained 21% of the variance in L2 reading comprehension scores.

When the order of entry was switched (Analysis A2), i.e., when L1 reading comprehension ability and L2 linguistic proficiency were entered in step 1 and pseudoword decoding proficiency in step 2, pseudoword decoding remained a significant predictor of L2 reading comprehension, explaining an additional 1.6% of the variation in reading comprehension.

However, when L2 real word decoding proficiency was entered in step 1 (along with L1 reading comprehension ability and L2 linguistic proficiency) (Analysis A3), L2 pseudoword decoding proficiency ceased to be a significant predictor of L2 reading comprehension. However, the fact that L2 real word decoding too ceased to make a significant contribution suggests that the non-significant contribution of L2 pseudoword is likely caused by the large correlation between the two (r=.77), resulting in inflated standard errors, and potentially, unreliable tests of significance.

Therefore, to avoid a violation of the assumption of collinearity[21], and to accurately test if the contribution of pseudoword decoding was moderated by the proficiency of the learner, L2 real word decoding proficiency was excluded from the analysis. As can be seen in analysis A4, no linguistic proficiency* pseudoword decoding interaction was found, b=–.006, t=–1.857, p >.05. It is worth noting, however, that the p value was approaching significance, p = .06, and that when the significance threshold was set at p<.10 (to allow PROCESS to generate the follow-up simple slope analysis), the pattern of the results suggested that, again, the contribution of L2 pseudoword decoding proficiency showed a tendency to decrease as linguistic proficiency increased.

[21] A model in which the two variables (L2 real word decoding and L2 pseudoword decoding) were included passed some collinearity tests (e.g., no correlation at r=.9), but not others (e.g., those related to variance proportions and conditioning index).

- **Analysis B:**

Table 33 below presents an overall summary of the findings of the second set of analyses (Analysis B), which aimed at: (1) examining the contribution of L2 real word decoding proficiency to L2 reading comprehension, before and after the contributions of L2 linguistic proficiency and L1 reading comprehension ability have been partialled out **(analysis B1 and B2)**; (2) examining whether real word decoding proficiency predicts any variance in L2 reading comprehension after additional controlling for L2 pseudoword decoding proficiency **(analysis B3)**; examining whether the contribution of L2 real word decoding proficiency would vary as a function of the linguistic proficiency of the learner, and if so, how **(analysis B4).**

Table 33. Summary of regression analyses for investigating the contribution of L2 real word decoding proficiency to L2 reading comprehension

Analysis B							
	B	**ß**	**T**	**R**	**R^2**	**Adj R^2**	**ΔR**
Analysis B1							
Step 1				.65	.42	.42	.42***
L2 real word decoding proficiency	.278	.647	12.298***				
Step 2				.79	.62	.61	.20***
L2 real word decoding proficiency	.083	.194	3.110**				
L1 reading comprehension ability	.213	.107	2.479*				
L2 linguistic proficiency	.411	.618	9.879***				
Analysis B2 (order of entry switched)							
Step 1				.78	.60	.60	.60***
L1 reading comprehension ability	.203	.102	2.319*				
L2 linguistic proficiency	.505	.760	17.314***				
Step 2				.79	.62	.61	.018**
L1 reading comprehension ability	.213	.107	2.479*				
L2 linguistic proficiency	.411	.618	9.879***				
L2 real word decoding proficiency	.083	.194	3.110**				
Analysis B3 (additional controlling for L2 pseudoword decoding proficiency)							
Step 1				.79	.62	.61	.62***
L1 reading comprehension ability	.195	.097	2.259*				
L2 linguistic proficiency	.462	.695	14.371***				
L2 pseudoword decoding proficiency	.073	.143	2.963**				
Step 2				.79	.62	.61	.00
L1 reading comprehension ability	.205	.103	2.379*				
L2 linguistic proficiency	.422	.635	9.836***				
L2 pseudoword decoding proficiency	.037	.073	1.046				
L2 real word decoding proficiency	.054	.126	1.397				
Analysis B4 (introducing the interaction term)							
Step 3				80	.65	.64	.03***
L1 reading comprehension ability	.161	.080	1.910				
L2 linguistic proficiency (centred)	.431	.648	10.644***				
L2 real word decoding proficiency (centred)	.057	.132	2.123*				
L2 RWDP*linguistic proficiency	-.011	-.174	-4.013***				
Note. $N = 212$; *$p < .05$, **$p < .01$, ***$p < .001$							

As can be seen in the table above (Analysis B1), from a hierarchical multiple regression analysis with reading comprehension as the criterion variable, and L2 real word decoding proficiency as a predictor entered at stage 1, a statistically significant model emerged (Adjusted R^2= .42, F =, p<.001), suggesting that real word decoding explained 42% of the variance in L2 reading comprehension scores.

When the order of entry was switched, i.e., when L1 reading comprehension ability and L2 linguistic proficiency were entered in step 1 and pseudoword decoding proficiency in step 2 (Analysis B2), real word decoding proficiency remained a significant predictor of L2 reading comprehension, explaining an additional 1.8% of the variation in reading comprehension.

However, when L2 pseudoword decoding proficiency was entered in step 1, along with L1 reading comprehension ability and L2 linguistic proficiency (Analysis B3), L2 real word decoding ceased to be a significant predictor of L2 reading comprehension. However, again, this non-significant contribution is likely caused by the high correlation between L2 pseudoword decoding and L2 real word decoding proficiency (r=.77). Thus, to test if the contribution of real word decoding proficiency to reading comprehension was moderated by the proficiency of the learner, L2 pseudoword decoding proficiency was excluded from analysis B4.

As can be seen in table 33 above, there was a significant proficiency* L2 real word decoding interaction, b=–.011, t=–4.013, p <.001. To visualize the nature of the moderation effect, the interaction was probed at high (+1 SD above the mean), average (mean) and low (–1 SD below the mean) levels of linguistic proficiency (Table 34 and Figure 37). When language proficiency was low, L2 real word decoding proficiency was a significant predictor of reading comprehension, with b= .135, t= 4.677, p < .001. When language proficiency was moderate, L2 real word decoding proficiency was also a significant predictor of reading comprehension, though the size of the contribution was lower, with b= .057, t= 2.123, p < .05. When language proficiency was higher, the contribution of L2 real word decoding to reading comprehension was non-significant, with b= -.022, t= -.595, p = .552. Together, these results suggest that the contribution of L2 real word decoding proficiency to L2 reading comprehension decreased as learners' linguistic proficiency increased.

Table 34. Summary of follow up simple slope analysis

Linguistic proficiency	B	t	p
Low (-7.032)	.135	4.677	.000
Average (.000)	.057	2.123	.035
High (7.032)	-.022	-.595	.552

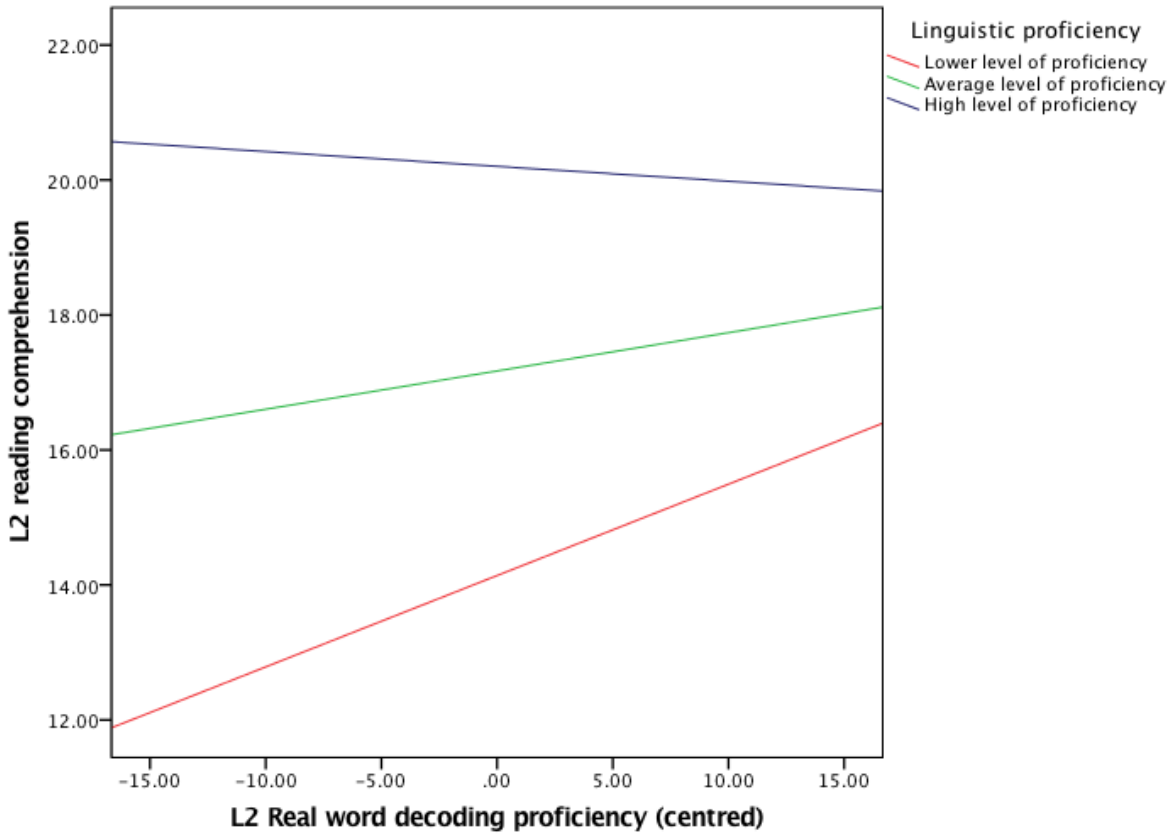


Figure 36. The contribution of L2 real word decoding proficiency as a function of the proficiency level of the learner

To conclude, both L2 pseudoword and L2 real word decoding proficiency made a significant contribution to L2 reading comprehension, both before and after the effects of L2 linguistic proficiency and L1 reading comprehension ability were controlled for. However, neither made an independent contribution to L2 reading comprehension after controlling for each other and the other measured covariates. Furthermore, the contribution of L2 real word and L2 pseudoword decoding proficiency seemed to decrease as linguistic proficiency increased, though the pseudoword*proficiency interaction was only approaching significance.

The fact that the two measures made a unique contribution to L2 reading comprehension after controlling for the other measured covariates, but not after additionally controlling for one another, suggest that these two measures are more related than different. Having said that, however, it is interesting to note that the composite measure of both L2 real word and L2 pseudoword decoding proficiency (the overall L2 decoding proficiency measure) made a slightly larger contribution to L2 reading comprehension (2.0%) (section 5.2.3), than the L2 real word decoding proficiency measure (1.8%) and the L2 pseudoword decoding proficiency measure (1.6%), again suggesting that the two measures, while clearly significantly related, remain different, albeit marginally.

5.2.5. The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance analysis and Relative weight analysis

The primary aim of this section was to determine the relative importance of L2 decoding proficiency to L2 reading comprehension using Dominance analysis and Relative Weight analysis. Furthermore, Braun and Oswald (2011) noted that it would be informative to examine whether (and how) different indices of relative importance converge and diverge. Therefore, a secondary aim in this section was to investigate whether the relative importance of decoding proficiency varies as a function of the analytic metric used with a particular emphasis on Dominance Weights, Relative Weights, and the more traditional Beta Weights.

The results of two sets of analyses are presented in this section. The first assesses the relative importance of *overall* L2 decoding proficiency to L2 reading comprehension, and the second assesses the relative importance of both individual measures of L2 decoding proficiency, (i.e., real word and pseudoword decoding proficiency). The rationale behind conducting these two sets of analyses is that the impact of collinearity in the second analysis may be greater, as the two decoding measures are highly correlated, presenting an interesting issue when determining importance. Indeed, one motivation for this analysis was the finding reported in section 5.2.4. Particularly, when the two decoding measures were included in one model, both emerged as nonsignificant despite the medium-to-large correlation they share with L2 reading comprehension, thereby highlighting the need for further examination of their relative importance to L2 reading comprehension. Dominance Analysis and Relative Weight Analysis were thus performed twice (i.e., for each set).

- The relative importance of overall L2 decoding proficiency to L2 reading comprehension

Table 35 below presents the Relative and Dominance Weights of four predictors: L2 Vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and overall L2 decoding proficiency. For comparative purposes, Beta Weights are also included.

Table 35. The relative importance of overall L2 decoding proficiency to L2 reading comprehension using Dominance, Relative, and Beta weights

Variable	Relative importance metric		
	Dominance Weights	Relative Weights	Beta weights
L2 Vocabulary knowledge	.225 (2)	0.222 (2)	.314 (2)
L2 Grammar knowledge	.265 (1)	0.263 (1)	.428 (1)
L1 Reading comprehension	.017 (4)	0.018 (4)	.108 (4)
L2 decoding proficiency	.131 (3)	0.134 (3)	.124 (3)
Note: *Between parentheses is the rank order of predictor importance.			

Four key observations can be made from the above table: (1) Dominance Weights and Relative Weights did not differ from Beta Weights in that, descriptively, the rank order of variable importance was consistent across the three metrics, with L2 grammar knowledge emerging as the strongest predictor of L2 reading comprehension, followed by L2 vocabulary knowledge, L2 decoding proficiency, and finally L1 reading comprehension ability; (2) Dominance Weights and Relative weights provided almost identical results to each other, which concords with the findings of previous simulation studies (Lebreton, Ployhart, & Ladd, 2004); (3) it is interesting to note the discrepancy between L2 grammar knowledge and L2 vocabulary knowledge across the three metrics. According to both Dominance Weights and Relative Weights, that discrepancy is ~15% (RW of .225 vs. .265, and DW of .222 vs. .263 for L2 vocabulary knowledge and L2 grammar knowledge respectively), which is half the discrepancy seen between their beta weights (.314 vs. 428, i.e., ~30%); and (4) again, the dominance and relative weights sum to the model R^2 (~.64), and thus can be used as a measure of effect size.

Unfortunately, no method has currently been reported to examine the statistical significance of Dominance weights (these are automatically provided for beta weights using any popular statistical software). However, RWA-Web conducts null hypothesis significance tests using advanced bootstrapping procedures to assess both (a) the statistical significance of the Relative Weights; and more importantly (b) the statistical significance of the *difference* between the Relative Weights generated (Tonidandel & LeBreton, 2015). These two types of information were deemed valuable and are reported below.

The relative weights of all four variables were statistically significant, as none of the 95% CIs contained zero, 95% CI= [0.177, 0.265], [0.213, 0.318], [0.001, 0.053], and [0.085, 0.184] for L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 decoding proficiency respectively. Further, it was possible to see whether the relative

importance of overall decoding proficiency was significantly different from the remaining reading comprehension variables. Results indicated that relative weight of L2 decoding proficiency (RW=0.13) was significantly lower than the weights of L2 grammar knowledge (RW= .26), 95% CI [0.05, 0.22] and L2 vocabulary knowledge (RW=0.22), 95% CI [0.02, 0.16], but significantly higher than that of L1 reading comprehension ability (RW= 0.02) 95% CI [-0.17, -0.06]. The difference between the relative weights of L2 grammar knowledge, L2 vocabulary knowledge, and L1 reading comprehension ability was also examined. As can be seen from table 44 (Appendix L), differences between all the weights were statistically significant except for that between L2 grammar knowledge and L2 vocabulary knowledge, as the 95% CI contained zero (Tonidandel & LeBreton, 2015), suggesting that the two variables may be equally important for L2 reading comprehension.

- The relative importance of L2 real word and L2 pseudoword decoding proficiency to L2 reading comprehension

Table 36 below presents the Relative and Dominance Weights of five predictors: L2 Vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 real word decoding proficiency and L2 pseudoword decoding proficiency. Again, for comparative purposes, beta weights are also presented.

Table 36. The relative importance of real word and pseudoword decoding proficiency to L2 reading comprehension using Dominance, Relative, and Beta weights

Variable	Relative importance metric		
	Dominance Weights	Relative Weights	Beta weights
L2 Vocabulary knowledge	0.197 (2)	0.200 (2)	0.321 (2)
L2 Grammar knowledge	0.236 (1)	0.250 (1)	0.435 (1)
L1 Reading comprehension	0.016 (5)	0.016 (5)	0.107 (3)
L2 real word decoding proficiency	0.129 (3)	0.116 (3)	0.084 (4)
L2 pseudoword decoding proficiency	0.059 (4)	0.060 (4)	0.039 (5)
Note: *Between parentheses is the rank order of variable importance.			

Four key observations can be made from the above table: (1) On a descriptive level, the rank order of variable importance differed as a function of the analytic metric used, with Dominance Weights and Relative Weights suggesting that L2 grammar knowledge is the strongest predictor of L2 reading comprehension, followed by L2 vocabulary knowledge, L2 real word decoding proficiency, L2 pseudoword decoding proficiency, and finally L1 reading comprehension ability. Beta weights, on the other hand, place L2 real word decoding proficiency and L2 pseudoword decoding proficiency last, after L1 reading comprehension ability, despite the much larger correlation between these two variables and L2 reading comprehension. (This is due to the correlations these two predictors share with one another, resulting in them explaining little unique variance); (2) Again, Dominance Weights and Relative weights provide almost identical results, which corresponds with findings of previous studies (Lebreton, Ployhart, & Ladd, 2004); (3) Though – descriptively – all three metrics suggest that L2 grammar knowledge is more important than L2 vocabulary knowledge, it is interesting to note the discrepancy between the weights of these predictors across the three metrics. According to both Dominance Weights and Relative Weights, that discrepancy is much smaller than that found between their beta weights; and (4) Again, the dominance and relative weights sum to the model R^2 (for example, the dominance weights 0.197+0.236+0.016+0.129+0.059= ~.64), and thus can be used as a measure of effect size.

Further, an examination of the relative weights reveals that all five variables explained a statistically significant amount of variance in reading comprehension, as none of the 95% CIs contained zero (corresponding to a significance testing alpha level of .05), 95% CI= [0.194, 0.297], [0.160, 0.242], [001, 0.052], [0.084, 0.150], and [0.030, 0.096], for L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, L2 real word decoding proficiency, and L2 pseudoword decoding proficiency, respectively. The relative weight results differ slightly from what was obtained from the traditional multiple regression analysis. Specifically, in the traditional analysis, neither real word decoding proficiency nor pseudoword decoding proficiency provided a statistically significant incremental effect in the prediction of reading comprehension, after controlling for one another (and all of the remaining reading comprehension variables). Consequently, the two predictors emerged as the least important to L2 reading comprehension as indicated by their descriptive rank ordering (fourth and fifth using beta weights vs. third and fourth using relative weights and dominance weights).

Additionally, results from the Relative Weight analysis suggests that the relative importance of real word decoding proficiency was significantly different from the remaining

reading comprehension variables. Particularly, the relative weight of real word decoding proficiency (RW=0.12) was significantly lower than the weights of L2 grammar knowledge (RW= .250), 95% CI [0.07, 0.20] and L2 vocabulary knowledge (RW=0.20), 95% CI [0.03, 0.14], but significantly higher than that of L1 reading comprehension ability (RW= 0.02) 95% CI [-0.14, -0.05] and L2 pseudoword decoding proficiency (RW= 0.06) 95% CI [-0.09, -0.02].

Similarly, the relative importance of pseudoword decoding proficiency was tested against that of the other predictors in the model. Results indicated that the relative weight of pseudoword decoding proficiency (RW=0.06) was significantly lower than that of L2 grammar knowledge (RW= .250), 95% CI [0.111, 0.256], L2 vocabulary knowledge (RW=0.20), 95% CI [0.080, 0.198], L2 real word decoding proficiency (RW= 0.12) 95% CI [0.020, 0.089], but significantly higher than that of L1 reading comprehension ability (RW= 0.11) 95% CI [-0.086, -0.003].

The difference between the relative weights of L2 grammar knowledge, L2 vocabulary knowledge, and L1 reading comprehension ability and the remaining predictors was also examined. As can be seen from table 45 (Appendix L), differences between all the weights were significant except for that between L2 grammar knowledge and L2 vocabulary knowledge, as the 95% CI contained zero.

To conclude, when the relative importance of overall L2 decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension was examined using Relative Weight Analysis and Dominance Analysis, the findings obtained were similar to those provided by the beta weights presented earlier (section 5.2.2) as, at a descriptive level, the rank order of variable importance was the same: L2 grammar knowledge, followed by L2 vocabulary knowledge, then overall L2 decoding proficiency, and finally L1 reading comprehension. However, results from the relative weight analysis suggested that the difference between the relative importance of grammar knowledge and vocabulary knowledge was not statistically significant, potentially suggesting that the two are equally important for L2 reading comprehension.

However, when the relative importance of L2 real word decoding proficiency, L2 pseudoword decoding proficiency, L2 vocabulary knowledge, L2 grammar knowledge and L1 reading comprehension was examined using Relative Weight Analysis and Dominance Analysis, the findings obtained produced findings slightly different to those of beta weights (section 5.2.4) as L2 decoding proficiency (pseudoword or real word) continued to appear more important to L2 reading comprehension than L1 reading comprehension ability suggesting that

Relative Weight and Dominance analyses were more able to detect the collinearity present within the data and to correct for these when determining variables' relative importance.

5.2.6. The contribution of second language decoding proficiency to second language reading comprehension: qualitative insights from readers' experiences

Analysis of participants' responses in relation to the issues interferring with their reading comprehension resulted in the emergence of eight categories, explored in turn below. For organizational purposes, categories are presented in order of their frequency of occurrence, with the exception of decoding knowledge which is discussed in detail last.

Categories
L2 vocabulary knowledge (100%)
Affective factors (44%)
Background knowledge (33%)
Grammar knowledge (28%)
Inferencing ability (22%)
Other
Decoding knowledge (50%)

5.2.6.1. L2 Vocabulary knowledge

Vocabulary knowledge – defined as word meaning knowledge in the present study – was the most frequent theme, occurring across all 18 interviews (i.e., in 100% of the sample).

For example, S7 mentioned that *"it's mainly my vocabulary [that interferes with my understanding]. I don't have much vocabulary*". S16 also noted that her vocabulary repertoire is composed of *"short, easy, and frequently used [vocabulary]"*, which in turn impacts her understanding of English texts. S18 also noted that her vocabulary knowledge is "not much" and that, as a result of this, she often finds herself having to infer the meanings of words from the sentential context, but that *"this inference turns out to be wrong, because I should have known the meanings of at least some words"*.

Comments also reflected issues with the depth of word meaning knowledge. For example, S4 mentioned that *"Sometimes, no not sometimes, always, I encounter words which I don't know, but I know their synonyms. I mean they don't use the familiar words, for example the words "complain" and "nag". In English texts, they use unfamiliar words."* A similar comment was made by S7: "*there are words that I know… but when I try to let them fit in the sentence, I feel the meaning is not matching with the rest of the sentence. The word would have more than one meaning*".

It is interesting to note that while the majority of the responses suggested that lack of L2 vocabulary knowledge interferes with one's ability to build overall meaning of the text, some students explained how vocabulary knowledge interfered with their understanding in other ways. For example, during the stimulated recall task, S11 mentioned that if she does not understand word meanings, she "*won't keep on reading*", suggesting that limited vocabulary knowledge could modulate situational affect, which might impact comprehension. Furthermore, S16 expressed how the presence of too many unknown words can be cognitively overwhelming:

> *Researcher: What's the problem with new words?*
> *S16: Their meanings... when the passage is too long. They fight in my mind.*
> *Researcher: Who fight?*
> *S16: The words. When you have many many words.*
> *Researcher: what do you mean by fight in my mind?*
> *S16: I mean more than one word I don't know and I am thinking of each: "What does this mean. What does that mean", like that.*

5.2.6.2. Affective factors

Eight of the 18 participants (i.e., 44% of the sample) referred to one or more affective factors. These include:

(a) anxiety (37.5% of the sample), as S6 comments:

"I am a person who gets nervous for no reason. And when I feel like 'I have to do this', 'I have to pass', I get so nervous. And it's not in your hand. It's uncontrollable. Stress will not let you understand quickly. You need time to feel calmer and understand."

(b) lack of interest in reading in general (25% of the sample), as the following exchange shows:

> *S2: "I don't like reading, be it in English or Arabic".*
> *Researcher: "why do you hate it in Arabic?"*
> *S2: "I get bored. I get bored so much. I find it strange that people read novels."*

(c) lack of interest in second language reading in particular (12% of the sample), as S8 noted: *"I don't like to read English much. I listen, yes, but read, no"*. When the researcher asked her if she feels the same towards Arabic reading, S8 commented, *"No, I'm obsessed about reading in Arabic"*, and said that she particularly dislikes reading in English because *"by the time I try to understand and comprehend, the joy of reading would be gone"*.

(d) lack of interest in the reading topic (25%), as S17 comments: *"Also, if it's a scientific text, I won't like reading it. Say in physics or history and like that. I won't like it. But if it's a story or a novel, I would love it"*.

5.2.6.3.Time pressure

As the sample comprised instructed second language learners whose L2 reading experience is often situated in timed testing settings, six of the 18 participants (33%) mentioned time pressure in their discussions of the issues that interfere with their understanding of English texts. For example, S2 noted that, when she is put under time pressure, she finds herself "*focus[ing] on the time itself and not on the task in hand"*. When she was asked if she would encounter a similar experience when reading for pleasure (for example, at home), she interrupted: *"No, time is not a problem here"*. S9 also expressed how time pressure can modulate her situational affect, *"If someone tells me: "now you have three minutes", I get nervous"*.

5.2.6.4. L2 Grammar knowledge

Five of the 18 participants (i.e., 28% of the sample) referred to issues with their grammar knowledge. These were often related to verb tense (i.e., present/past/future) or voice (active/passive), as the following two exchanges show:

> *S9: sometimes the grammar they use messes me up.*
> *Researcher: How does it interfere with your understanding, could you explain?*
> *S9: like for example when two tenses are mixed. Here I struggle…like for example if he [the writer] was using the future and also the present. Here I struggle. I am like, you decide, is this now or in the future?*
>
> *Researcher: So you said the vocabulary in this text was difficult. So when you know the meanings of all of these words, you would be able to understand it well?*
> *S18: maybe, if I know the meanings, but not all all of it.*
> *Researcher: why not all of it?*
> *S18: because vocabulary is not everything from my experience… some grammar points I face difficulty in understanding them. Some grammar points I have not even grasped them. I mean I have not studied them yet and I don't understand them. For example, like the passive and this kind of thing…It just did not enter my mind. I don't know. I find myself helpless when I want to understand it. I just could not understand it. And also that of the gone and been I don't understand it. But I do understand some, like when they use the present simple and things like these. The past.*

5.2.6.5. Background/topic knowledge

Four of the 18 participants (i.e., 22% of the sample) referred to background knowledge as one of the factors that contribute to their reading comprehension performance. To exemplify, two excerpts are provided below:

> *Researcher: and if you understood the meanings of all of these words? Would you be able to understand the passage?*
> *S18: roughly*
> *Researcher: why roughly? What's left?*
> *S18: the passage could be about something I don't know. That's why I won't connect to it much. About a topic I don't know. Like a scientific topic or something I don't know.*

> *S5: If the topics were complex*
> *Researcher: like what?*
> *S5: like the scientific subjects that I am taking now. The vocabulary is not the vocabulary that we've always known. Scientific terms. A bit complex. This interferes with my understanding too.*

5.2.6.6. Inferencing ability

Four of the 18 participants (i.e., 22% of the sample) referred to inferencing ability when discussing the issues that interfere with their understanding of English texts. For example, S6 expressed how in the past (i.e., when she was in school), she was used to encountering mostly literal comprehension questions which she found manageable to answer: *"when the question is there and you just look for the exact words that make up the question in the passage and then answer automatically. They don't change the question ... No it's the same. Copy and paste"*. She then explained how when she started university, she started seeing difficult and "*weird*" questions that force her to infer meanings not explicitly stated in the text: *"I notice they do that a lot here, when they present the passage and then ask you to choose a title for this passage"*. A similar comment was made by S9 during the stimulated recall task: *"In questions like these (referring to the text), if the meaning is literal I'd be able to find the answer, but if you want the real meaning, I won't be able to answer*".

5.2.6.7.Other

Smaller number of participants referred to factors other than the ones described above (and decoding below). One of these factors was concentration issues (mentioned by 2 participants, i.e., 11% of the sample), as S4 noted, "*When I read, sometimes, I get very immersed in the topic and think: "why did he use this word and not that" and end up leaving the topic that I am reading about*".

Another factor was the reading atmosphere, raised by one participant (i.e., 6% of the sample): *"Noise. The surrounding environment. Particularly when it's English that I have to read"*.

A third factor also raised by one participant was text length, as S6 explains:

"Also, when the passage is too long. It makes things difficult. I have just learned this and I would need to learn that. It would be tiring… It is tiring to memorize. Like you have to understand this and memorize it and move to another thing and the same thing over and over".

Another factor subsumed in this category was knowledge of derivational morphology, which was inferred from one participant in the exchange below, which took place during the task-based discussions:

S4: there are other words I don't know.
Researcher: like what
S4: richness
Researcher: Could you not guess its meaning?
S4: (thinking)
Researcher: if you divide it you might know
S4: Ah rich plus ness.

Thus, these "other" factors were mentioned by a relatively smaller number of participants, but were nonetheless relevant to those who raised them.

5.2.6.8. L2 Decoding knowledge

Nine of the 18 participants (i.e., 50% of the sample) referred to decoding knowledge when discussing the issues affecting their understanding of English texts. It is worth noting here that seven out of these nine participants (i.e., 78%) were classified as A1 (N=4) and A2 (N=3) level learners.

On several occasions, students mentioned that "long words" interfere with their comprehension. To ensure that they were referring to a decoding issue, rather than a vocabulary one, the researcher often asked follow-up questions, such as, 'What is the problem with long words?'. Only where the response reflected a decoding issue was it coded as "decoding".

In what follows, I report the mechanisms inferred from the responses of the participants who mentioned decoding as one of the factors contributing to their reading comprehension performance. On occasions, students would have already explained the mechanism as part of their response to the general question (i.e., 'What things interfere with your understanding of English texts?'). However, there were instances where the participant did not provide further details as to 'how' decoding knowledge interfered with their understanding, which prompted

the researcher to ask the follow up question, "How?". Additionally, there were instances when the mechanism was inferred from student-researcher discussions during the stimulated recall task. Thus, the mechanisms reported below were derived from (a) students' initial response to the above interview question, (b) students' response to the follow up question ("How?"), and (c) student-researcher discussions during the stimulated-recall task.

- ***The meaning access mechanism***

Meaning access was the mechanism which emerged from the data of 8 participants. For example, S3 eloquently explained how silent letters affect her decoding performance and her subsequent understanding of the word's meaning:

> *When letters are silent and I don't know they are silent. They affect my reading of the word and my understanding of it. Sometimes I know a word. I listen to it. I know what it means but I don't know how it's written. So when I see it, and I see different letters, I don't know how to read it so I don't know its meaning, and I know its meaning, do you understand me?*

S4 and S6 also explained how they can understand the meaning of a word upon hearing its spoken form, but less so upon seeing the word's written form suggesting lack of decoding knowledge:

> *S4: Also, words which I have heard but never seen. I have never seen the way they are written. What their spelling is. When someone speaks to me, I might listen to it and know what he means, but when I read it, it doesn't it doesn't come to my mind. I read it like it's a new word to me."*

> *S6: There are words which it's not that you don't understand their meanings. Sometimes a word [word meaning] is easy. I read a text [but do not understand the meaning of the word] and a student reads it out loud and I say this is a known word, but I could not read it on my own.*

The stimulated recall data was a useful source of insight into the mechanisms underlying the contribution of decoding to reading comprehension, and particularly, the meaning access mechanism. As can be seen below, it is clear that the student understands the meaning of the word "Casablanca". In fact, "Casablanca" is commonly used in spoken Arabic, also to refer to the city. However, the student's inaccurate decoding has interfered with her understanding of the meaning of the word, thus potentially disturbing her overall reading comprehension:

> *S3: this word (pointing at the word Casablanca). I don't know its meaning and I don't know how it's pronounced.*
> *Researcher: try to read it*

S3: /Kæs/. Does it start with K?
Researcher: Yes. How did you know?
S3: I don't know I just guess. I don't know. /Kæs/… (tries to read the rest of the word) /blæns/. /Kæsblæns/. Is this correct?
Researcher: /kasəblankə/ *(the researcher provides the accurate pronunciation)*
S3: Aha Casablanca the place?
Researcher: Yes.

A similar example is also provided below. The student pointed at the word "control" as a difficult word in the passage (again, <control> is frequently used in spoken Arabic)

S10: this (pointing at the word control)
Researcher: whats difficult in this
S10: /kʌn/… /tu:/… /rɒl/. Like that maybe? I don't know.
Researcher: what's difficult in it?
S10: made me slow. I don't know. The letters made me pause
Researcher: do you know its meaning? Control? (The researcher provides the accurate pronunciation)
S10: control? (the student reads it again correctly following the researcher's pronunciation and provides the accurate meaning)
Researcher: yes
S10: Ahaa. This (i.e., pointing at the word) means control?
Researcher: yes. Why didn't you get the meaning when you read it yourself?
S10: that's not how control is written. I thought control is written differently.

On a side note, it is worth mentioning that a similar exchange was found in discussions with a participant who did not take part in the present interview but who participated in the decoding experiences interview. The encounter is presented below for the interested reader:

Participant: like this word (points at the word mosque). I thought it's 'Moscow', turned out it is 'mosque', 'masjid' (masjid is the Arabic translation of mosque).
Researcher: and what does Moscow mean?
Participant: the city

Further examples were found in participant responses to direct follow-up questions about decoding:

S9: Also, if there were silent letters. Those that I don't say.
Researcher: and how does this interfere with your understanding?
S9: I can't pronounce them. I worry that I pronounce them and then the letter should not have been pronounced. Like that. I don't know the pronunciation of the word. I worry. I know the word, and I would have listened to it before, but the letters are not pronounced. If I don't pronounce them I would know the meaning.

S10: Also the words. The letters. Some letters like S H. In some letters [words] I know they would be S plus H equals sha like that but some like T and U or... I think T and U. I don't know but it's also <sha>. And when I read them it comes wrong because they would be pronounced using other letters.
Researcher: But how does this interfere with your understanding?
S10: when there is a word that I know, for example like.. So a word that I know, but don't know how it's written and I read it, I read it incorrectly
Researcher: how did you know it then? The word.
S10: I don't know it. I know it. I know it from from my experience. When they talk and I memorised it.
Researcher: when who talks?
S10: in school, in a movie I watched, at home, outside, everywhere.

To conclude, all the examples presented above suggest that decoding proficiency contributes to these learners' reading comprehension through its role in facilitating accurate meaning access.

- ***The affective mechanism***

Three of the nine participants (i.e., 33% of the sample) expressed how encountering difficult-to-read words impacted on their affective state.

For example, S12 and S11 both mentioned that "long words" interfere with their understanding. When asked "what is the problem with long words?", S12 explained how their presence might trigger anxious feelings, leading to feelings of confusion: "*When the words are long, and also I have limited time, the moment I see them long, I get nervous. I confuse things".* In her response to the same question, S11 explained how the presence of long words trigger feelings of self-doubt: *"Like, when you see them, you feel like, 'No, I can't read them', which means [their] meaning would be difficult, you would not know [them]". This is despite the fact that* she might be able to decode the word if she attempted to do so: "*But when you try to read it, it might work with you, but some words, no, it won't. Because you don't know them completely".*

A further example is provided below. When S5 was asked how her decoding interferes with her understanding, she explained how her inability to read words triggers feelings of disappointment, and how these feelings interfere with her understanding.

Researcher: and you said that you don't pronounce words correctly, but I just want to understand more how does this interfere with your understanding. Could you explain?
S5: I feel disappointed. I feel like I'm not made for English. Although I swear I love English.
Researcher: and when you get these feelings, how do you find your understanding? For

example, do they make it better, worse, or they don't do anything really?
S5: No, make my understanding worse. It makes me feel, 'Oh, I don't know this word'.

Finally, although only 3 of the participants spontaneously mentioned how decoding proficiency could impact their affective state, it is worth noting that the researcher decided to ask all participants whether they experienced any feelings when encountering difficult-to-read English words. Where they responded yes, they were further asked what these feelings were, and whether experiencing them facilitates, worsens, or does not interfere with their understanding of written texts. 83% of the participants noted that they do experience feelings when encountering difficult-to-read words (whereas 17% noted that they do not encounter any such feelings). When asked what these feelings are, disappointment was the most-reported feeling (47%), followed by anxiety (27%), and 'other' (47%), which included: engaging in negative self-talk; sadness; and irritability. When asked how they find their understanding when encountering these feelings, 8 participants (53%) noted that they worsen their understanding, one participant (7%) noted that they improve their understanding, and one participant (7%) noted that they do not interfere with their understanding. The responses of five participants (33%) were categorized as "other", as they did not easily fall into these categories (These responses were along the lines of: "*sometimes these are temporary feelings and sometimes not*", "*I say to myself I have to try*", "*I push myself*", "*bad but sometimes it pushes me to learn*").

- ***The cognitive overload mechanism?***

No participant spontaneously expressed how decoding difficulties might overload their mental capacity. However, when asked about their comprehension experiences when encountering words that take them some time to read, 15 of the participants (80% of the sample) noted that their decoding speed interferes with their understanding. Below are two examples:

> *S11: when words are short, and I read fast, my understanding would be fine. But when a word is short, it would interrupt my understanding.*
> *Researcher: when it is what?*
> *S11: ah, when it's long, it interrupts my understanding*
> *Researcher: how does it do so?*
> *S11: like, I am going well, and really immersed in reading and then I see this word and I feel like because I think about it too much, I forgot what I had read before.*

Researcher: and how is your decoding speed?
S3: it depends on the word. If it's a word I am used to writing it, reading it, using it in sentences, my speed would be good.
Researcher: like which words?
S3: "going", "back", "time", "old".
Researcher: and words that take you time to read?
S3: words that are difficult. Which I have not seen before. The long ones. Longer than five or four letters….
Researcher: how is your understanding when encountering these?
S3: my understanding becomes slow and broken into parts. Since I'm reading slowly, I already forgot the idea above. When I'm spelling out words, and thinking that's a "g" but should be pronounced "j", that's a U or what, I would have already forgot the idea before.

To conclude this section, the thematic analysis of the interview data suggests that learners' decoding proficiency could contribute to their reading comprehension performance, with this being due to the role decoding plays in facilitating meaning access, modulating situational affect, and, possibly, freeing cognitive capacity. Furthermore, at a descriptive level, the reading comprehension performance of lower-level learners appeared to be more disrupted by their decoding difficulties than higher-level learners.

Other factors that emerged as contributing to reading comprehension include vocabulary knowledge, affective state, grammar knowledge, time pressure, background/topic knowledge, inferencing ability, reading atmosphere, morphological knowledge, text length, and concentration issues.

6. Discussion

This chapter is divided into two sections. The first section (section 6.1) discusses the findings related to the decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students, whereas the second section (section 6.2) discusses the findings related to the contribution of second language decoding proficiency to second language reading comprehension. In both sections, findings are first discussed in light of previous theoretical and empirical literature. The implications of these findings to theories of reading and language pedagogy are then discussed. The sections conclude with a discussion of possible limitations of the present study and recommendations for future research.

6.1. The decoding abilities, realisations, experiences, and processes of Saudi foundation year students

6.1.1. Decoding abilities:

6.1.1.1. The findings on Saudi L2 learners' decoding abilities and the literature

The first research question sought to determine whether, and the extent to which, Saudi L2 learners (CEFR: A1, A2, and B1) encounter difficulties decoding their L2, English. Consistent with the literature on L2 learners in general and Saudi learners in particular (e.g., Hamada and Koda, 2008; Alsubaie, 2014), participants' performance on the decoding measures suggest that they do indeed encounter decoding difficulties.

Furthermore, some patterns were also observed. For example, learners appeared to encounter more difficulties reading pseudowords than real words, which corresponds with previous findings in the L2 domain (e.g., Woore, 2016), and with word reading theories (Coltheart, 2005; Seidenberg & McClelland, 1989). To illustrate, according to connectionist theories (e.g., Seidenberg and McClelland, 1989), familiarity with a certain grapheme-phoneme unit means that the links connecting the two (i.e., the graphemes and their corresponding phonemes) are 'solid'. Therefore, a reader is less likely to experience decoding difficulties, whether in the form of decoding accuracy or fluency, when reading real words (particularly familiar ones) as opposed to pseudowords. Similarly, dual-route models (Coltheart, 2005) assume that the "indirect" route (which is presumably employed when reading unfamiliar words) is less efficient than the direct route (which is presumably employed when reading real and familiar words). This is because employing the indirect route is assumed to result in a slower and possibly inaccurate decoding, as learners will need to compute

pronunciations, rather than retrieve them from memory, and that computation process may not be executed successfully, particularly in the absence of L2 GPC knowledge.

Furthermore, the cross-sectional data shows another finding of interest. First, for each of the decoding measures, significant differences were found between the three proficiency groups. Thus, the present data tentatively provides evidence for a linear development of decoding skill, at least for learners at the lower end of the proficiency scale. Needless to say, however, this finding should be viewed with caution given that the design of the study was not longitudinal. This linear pattern of growth corresponds with data from the L1 reading context (Torgesen, Rashotte, & Wagner, 2012), though it should be noted that, in the case of the L1 English reader, decoding development eventually levels off as decoding accuracy reaches ceiling or near-ceiling level (García & Cain, 2014).

6.1.1.2.Implications for theories of reading

At the outset, it should be noted that the aim of this question was to determine whether learners encounter difficulties decoding the second language, and as such, significant contributions to theory were not expected here. However, the findings provided some interesting theoretical insights which are discussed below.

To begin with, the SLA field witnesses a growing interest in diagnosing and outlining the linguistic abilities of learners at each stage of L2 development. For example, in the field of L2 vocabulary acquisition, Milton (2010) sought to identify the vocabulary size of learners at each stage of L2 development (with "stage" being operationalized as learners' assigned CEFR level). This line of research is valuable for theory of L2 development, as it establishes the developmental milestones for each language skill, thus enabling a clearer understanding of the linguistic abilities typically expected at each stage of development. As far as decoding ability is concerned, however, the picture is quite vague. Macaro and Erler (2011) noted that it is unclear as to the stage at which L2 learners acquire the GPC of the second language (Interestingly, this very question implies that (a) L2 decoding proficiency develops throughout the course of L2 development; and that (b) L2 learners will, at a certain stage, entirely master L2 decoding skill). This lack of clarity regarding whether, and when, L2 learners will acquire L2 decoding skill is unsurprising, as the field has yet to identify the scale of decoding knowledge typical at each stage of L2 development. The current study presented the beginning of such an investigation by quantifying the decoding success of learners at three developmental stages, specifically, the three lower levels of the CEFR hierarchy (A1, A2, and B1). Findings from these learners suggest that L2 decoding knowledge may increase roughly at a similar rate

throughout these lower developmental stages, though it remains unclear as to whether there will be a stage where decoding mastery is achieved, and if so, at which stage, and under what conditions.

6.1.1.3.Implications for language pedagogy

Given the popularity of the communicative language learning approach, minimal attention is given to developing decoding skill in L2 contexts (Macaro, 2007), including the Saudi one (Alsubaie, 2014). However, learners' low performance on the decoding measures indicates lack of decoding skill. One way that may improve learners' decoding skill is to introduce phonics instruction in language classrooms. In fact, prior intervention research showed positive improvements in decoding skill following a program of phonics instruction both in the L1 (Ehri, Nunes, Stahl, & Willows, 2001) and L2 (Li, 2019; Woore, Graham, Porter, Courtney, & Savory, 2018) contexts.

One might argue that, in the case of L2 English, it may be difficult to teach phonics due to the deep nature of the writing system. However, as Woodcock (1987) noted, approximately 80% of English words can be considered regular. Furthermore, as Share (1995) argued, there is some level of regularity even among words categorized as "exceptions". One might also argue that there are too many grapheme-phoneme patterns in English to be taught explicitly (hundreds as estimated by Gough and Hillinger (1980)) . However, it has been suggested that teachers could begin by teaching basic grapheme-phoneme relations, and following from that, readers may continue to learn further grapheme-phoneme relations implicitly (Ehri, 2015). Given the potential impact of implicit learning on further development of decoding skill, teachers could encourage L2 readers to engage in inductive learning (which creates opportunities for self-teaching, hence the description of decoding as a self-teaching mechanism (Share, 1995, 2008, 2011)). It is worth noting here that, in the L1 context, this self-learning process is facilitated by the fact that L1 readers are likely to have a large store of oral vocabulary (at least in comparison to L2 readers), which they can reference when attempting to decode a word. As such, they could match their decoding output to the pronunciation of an orally-acquired word, and this in itself creates an opportunity to enhance their GPC knowledge. As L2 learners are expected to be less exposed to auditory L2 vocabulary, teachers could encourage students to (a) increase their exposure to oral English and (b) to attempt to listen to the pronunciations of novel written words (say by using online dictionaries which often provide word pronunciation). This in turn would create opportunities for self-learning and eventually enhance learners' L2 GPC knowledge.

Furterhmore, it should be noted that while the goal of phonics instruction is to equip learners with the necessary decoding knowledge, teachers should aim to help L2 readers transition to a stage where they read words fluently. This transition could possibly be achieved by increasing learners' exposure to print. To date, it is unclear how many times a reader would need to encounter a novel word for it to be eventually automatically decoded. Hoover and Gough (1990) argue that one encounter is sufficient, though to the researcher's knowledge, this proposition is yet to be empirically tested.

Finally, it may be worth commenting on the implications of these findings to students in EMI contexts. If A1, A2, and B1 learners are studying at degree level, yet encounter decoding difficulties (as shown by their low decoding scores, and by the observation that these scores correspond to those of grades 1-3 children), then their decoding skill could arguably impact their learning of university level materials. If this is the case, then teachers would need to be made aware of the decoding difficulties of their students and to take steps to mitigate any problems they may cause (e.g. by making explicit links between written/spoken words). As mentioned above, introducing phonics teaching may be useful. This teaching could take place in ESL courses, remedial courses at HE level, or indeed in schools, so that students arrive onto EMI courses with better decoding proficiency. Another implication might be that students are tested not only on their overall English proficiency on entry to HE, but also on their decoding proficiency, so that institutions can monitor the scale and nature of the problem and thus plan remedial action as necessary.

6.1.1.4.Limitations of the findings

Perhaps three limitations are worth noting in this section. First, the present study examined the decoding abilities of A1, A2, and B1 level learners. Thus, it is unclear whether, and the extent to which, learners at the higher end of the proficiency scale would encounter decoding difficulties. Second, phonics instruction is not implemented at the data collection site (nor in Saudi Arabia in general, (Alsubaie 2014)); thus, L2 readers in other contexts who receive systematic phonics instruction may exhibit different decoding abilities. Third, the present study recruited learners with Arabic backgrounds. Given evidence for the impact of the L1WS on L2 decoding, it is possible that A1, A2, and B1 learners from other L1WS backgrounds would exhibit different decoding abilities (for example, learners of L2 English with alphabetic L1WS backgrounds may exhibit higher decoding abilities).

6.1.2. Decoding realisations and experiences

6.1.2.1. Findings on learners' decoding realisations and experiences

6.1.2.1.1. Decoding realisations:

Error analysis of learners' decoding outputs showed difficulties with decoding both vowels and consonants, a tendency to add to, omit, and mis-order words' phonemes, along with a tendency to make whole-word substitutions. Most of these categories of errors accord with earlier studies involving Saudi and Arab L2 readers (e.g., Alsubaie 2014; Emrey, 2005). However, in-depth analysis of the nature of these realisations provided additional insights.

With respect to the vowel realisations, error analysis suggests that there was little evidence of "randomness" in the realisations observed. As discussed previously, this was shown in (1) the unequal distribution of the vowel realisations obtained for any given vowel phoneme, with some realisations occurring significantly more frequently than others; (2) where a vowel phoneme had more than one graphemic representation, the observed realisations, and/or their proportional frequencies, tended to differ as a function of the graphemic context; (3) the three most frequent realisations observed for each vowel phoneme – which accounted for an average of 84% of the total short vowel realisations – seemed to differ as a function of the target vowel. These findings tentatively question the vowel blindness hypothesis put forward by Ryan and Meara (1996) and resonate with the findings of Alhazmi, Milton, and Johnston's (2019) study, which found no differences between the eye movement patterns of Arabic and non-Arabic readers (i.e., that both groups paid equal attention to vowels and consonants while reading). Second, it was interesting that some (though not all) of the vowel realisations observed tended to (a) share a considerable degree of articulatory features with the target vowel (i.e., in terms of vowel height, backness, and lip rounding); (b) reflect "regularization" errors (e.g., decoding <child> as /tʃild/); and or (c) tendency to pronounce vowel letter names (e.g., the second most frequent realisation of target <i> was /ai/). These findings not only tentatively question the "vowel blindness" hypothesis, but also suggests that vowel reading errors were potentially, though not purely, the outcome of (1) lack of L2 grapheme-phoneme knowledge; and (2) articulatory difficulties.

With respect to consonants, it was interesting that incorrect realisations of consonants (with one-to-one GPC) (specifically, <p>, <b>, <d>, <ch>, <sh>, <f>, <v>, and <j>) often shared the same place of articulation of the target phoneme (e.g., decoding <vasp> as /fa:sp/). That is, both /p/ and /b/ are bilabial sounds, /v/ and /f/ labio-dental, /tʃ/, /ʃ/, and /dʒ/ and /ʒ/ are post alveolar. As almost all of the problematic consonant sounds do not exist in the learners' L1

(particularly, /p/, /v/, /tʃ/ and /dʒ/) these consonant errors appear to be largely the outcome of articulatory difficulties, which supports the observations of Alsubaie (2014). However, one exception here was the grapheme <d>, as the most frequent incorrect realisation observed for this consonant was the bilabial /b/. (Note that also for both <p> and <b>, alveolar /d/ emerged as one of the frequent realisations). This suggests that these errors were potentially the outcome of a lack of familiarity with the L2 script as learners appeared to confused letters with similar shapes.

Finally, it was also interesting to note that for consonant graphemes with more than one phoneme correspondence, (for example, /s/ and /k/ for <c>) the incorrect realisations almost always pertained to the use of a regular but incorrect phoneme. For example, all errors decoding <c> in the pseudoword <clirt> pertained to the production of /s/, /slirt/, which is a regular phoneme for <c>. This could suggest that for consonants with less reliable GPC, consonant reading difficulties were mostly the outcome of lack of L2 GPC knowledge. Learners may potentially have inaccurate assumptions about the "shallowness" of the English writing system (i.e., they may rely on small, rather than large grain sizes when decoding consonants with less reliable GPC).

As for the phoneme insertion, addition, and misordering cases, the errors observed yielded several insights: First, the fact that phoneme insertion was observed in the silent letter context and the consonant cluster contexts suggest that these errors were mostly the outcome of (a) lack of L2 GPC knowledge (in the case of the silent letters), and (b) articulatory difficulties (in the case of consonant clusters), especially because the the clusters observed do not exist in Arabic (Smith, 2001). Second, the finding that phoneme omission errors were mostly observed in the consonant cluster (e.g., decoding <skree> as /sikri:/) and in the final <ee> context (e.g., decoding <shlee> as /ʃel/) again suggests that these errors could possibly be caused by articulatory difficulties, in the case of the former, and lack of L2 GPC knowledge, in the case of the latter (i.e., assuming that <ee> in the final position is silent). Third, the finding that phoneme misordering errors were mostly observed in the consonant cluster context again suggests that these could be the outcome of articulartory difficulties. However, it is important to note here that it was not possible to identify the patterns surrounding 17%, 64%, and 68% of the phoneme insertion, omission, and misordering cases, respectively. This means that there may be other explanations for this decoding behavior aside from lack of L2 GPC knowledge and articulatory difficulties. These could include inefficient processing of words and/or difficulties with phoneme blending, as Emrey (2005, p.138) explained with respect to the latter,

"If a subject is struggling to alphabetically decode a word, i.e., to read it letter by letter, she may easily forget which letters she has just pronounced and therefore struggles to blend them with the following letter-sounds".

Furthermore, the present study additionally showed that, at a macro level, no *qualitative* differences appear to exist between the different proficiency groups (A1, A2, B1) in the nature of their decoding difficulties as, at least at a macro level, the same main categories of error were observed in each group of proficiency. Nonetheless, *quantitative* variations between the groups were observed, as the error rates in each of the emerging categories appeared to decrease as proficiency level increased (though they were often more apparent between the A1 and B1 groups than between the A1 and A2 or A2 and B1 groups).

6.1.2.1.2. Decoding experiences

Thematic analysis of the interview data seemed to support findings from the analysis of decoding outputs, in that learners explicitly commented on their lack of L2 GPC knowledge (of both consonants, vowels, and silent letters) and on difficulties with the script of the L2WS. However, in addition to these, learners also reported difficulties decoding "unfamiliar" and "long" words. Difficulties reading unfamiliar words suggest that the learners encountered problems "computing" the phonological forms of words. It is unclear why long words were found difficult, but it could reflect blending problems (Ehri, 2015), an issue that was also reported in Alsubaie's (2014) study.

6.1.2.2.Implications for theories of reading

Findings from learners' decoding outputs and experiences have two implications for L2 decoding theory, and particularly, for Koda's Orthographic Distance hypothesis (2013). First, the fact that students clearly encountered difficulties with mirrored letters (e.g., d and b), and with articulating phonemes that do not exist in their L1, suggest that an L2 decoding theory should not only discuss writing system distance at the level of the unit of representation and orthographic depth, but also at the level of phonemic inventories and script. In other words, similarity in script and phonological inventories could potentially accelerate L2 decoding development.

Second, the lack of evidence supporting "vowel blindness" could suggest that the vowel decoding proficiency of Arabic learners may not be necessarily impeded by their consonantal processing experience. As the researcher has argued previously (section 2.2.5.1.1.2), this is not unexpected given that they are also literate in an alphabetic writing system, i.e., short-vowelled Arabic (nonetheless, this processing experience is often overlooked in discussions of Arab ESL

readers (section 2.2.5.1.1.2)). Thus, Koda's (2013) framework should account for situations were learners are exposed to more than one L1WS, as is the case of Arabic ESL readers who are biliterate in two writing systems: alphabetic (i.e., short vowelled Arabic) and consonantal (i.e., non short vowelled Arabic).

6.1.2.3.Implications for language pedagogy

It has been mentioned earlier that introducing phonics instruction may contribute to improving the decoding abilities of Saudi L2 learners. Findings from the decoding outputs and experiences data provide further insights for pedagogy as they show us what areas are particularly worthy of instructional attention among these learners. For example, students appeared to lack knowledge of basic grapheme-phoneme mapping rules. These include but are not limited to: (a) the "magic" e rule (i.e., when <e> is the final letter in a word that has one vowel phoneme, e.g., the split diagraph <a-e>, as in <late>, the <e> is usually silent and the sound of the preceding vowel changes); (b) the soft and hard <c> rule (i.e., when <c> is followed by the vowels <e>, <i>, or <y>, as in <cent> it will usually be decoded as /s/, and with other vowels, for example, <a> in <cat>, it would usually be decoded as /k/); (c) the soft and hard <g> rule (i.e., that <g> is usually decoded as /dʒ/ when followed by <e>, <i>, or <y>, as in <gym>, and as /g/ with other vowels, as in <yogurt>); (d) that repeated consonants in English are not geminated; and so on. Teachers could equip students with these "rules of thumb", though they should also raise students' awareness of the depth of the English writing system, as there may be exceptions to these rules.

Furthermore, teachers could prepare exercises geared towards improving learners' ability to differentiate between mirrored letters (for example, <b>, and <d>) so as to help them overcome any script-related difficulties. Also, teachers could familiarise students with multi-letter spelling units, as this may help them blend phonemes more easily, which would particularly aid the decoding of long words. Additionally, the teaching of letter names could be confusing for learners as they may "form a misconception that the letter names (rather than their phonemic values) form the basis for sounding out L2 words" (Woore, 2011, p. 395). Thus, teachers should also inform learners of the limitations of learning alphabetic letter names (ibid). Furthermore, teachers could prepare "phoneme blending" exercises to improve learners blending performance. Finally, learners clearly encountered difficulties articulating sounds that do not exist in their L1. To date, it is not clear as to whether these "articulatory difficulties" have any impact on their reading comprehension (or other areas, for example, speech intelligibility). However, some data suggests that this might be the case. As mentioned previously, in a study

where Arabic participants were asked to translate English words to their mother tongue, some of the translation errors observed appeared to reflect articulatory problems, for example, confusing <pillow> with <blow> or <invent> with <infant> (Emrey, 2005) (note that in these examples, confusion seemed to not only arise from articulatory difficulties as for example, the phoneme /i/ was also omitted in the <pillow, blow> case. However, it is possible that, in addition to other decoding difficulties, articulatory problems further contributed to altering the lexical identity of the word). If these articulatory difficulties prove to be problematic, then it may be useful to introduce English from a young age so as to facilitate the acquisition of its phonemic inventory. Finally, the finding that the error categories were observed in A1, A2, and B1 learners suggests that efforts need to be made to improve the decoding proficiency of learners at all three levels of proficiency (though to differing degrees, as lower proficiency learners encountered more decoding difficulties than higher proficiency learners).

6.1.2.4. Limitations of the findings

As noted previously, one-limitation of error-analysis is that learners may avoid reading words which they find difficult, which could potentially lead to an incomplete understanding of learners' decoding difficulties. Although this limitation was somewhat mitigated by investigating learners' decoding experiences, it is still possible that this avoidance behaviour concealed some information about learners' decoding difficulties. Furthermore, the small sample size in the between-groups analysis (20 per group) made it possible to see whether the main categories of errors were observed across groups. However, it would be interesting to run the between-group analysis on the sub-categories observed (for example, difficulties with silent letters and mirrored letters). Unfortunately, due to the small sample size, it was not possible to achieve this in the present thesis. Finally, it may be difficult to generalise the findings of the decoding experiences to the larger population of Saudi ESL learners, given the small number of participants taking part in the interview.

6.1.3. Decoding processes

6.1.3.1.The findings on learners' decoding processes and the literature

Analysis of learners' think-aloud protocols yielded three interesting findings, each of which is discussed below.

First, at a macro-level, the findings suggested that learners engaged in a range of decoding processes including (1) automatic processing; (2) assembled processing, evident in small-unit processing, large-unit processing, conscious computing of words' phonology by means of (a)

referencing L2 GPC knowledge, (b) making analogies (analogy-guided processing), and/or (c) using intuitions (intuition-guided processing); and also (3) partial/incomplete processing. Generally speaking, these processes seemed to support the dual-route framework, as learners appeared to employ the "direct" and "indirect" route to read the task items; the former being evident in the "automatic processing" category, whereas the latter was evident in all the categories underlying "assembled processing"[22]. The findings, however, do not support the 'original' conception of the dual route model, where the two processing routes were perceived as completely independent of one another. This is because, as was shown earlier, the processes observed did not seem to operate exclusively from one another as, on occasions, they cooperatively contributed to the production of the final phonological form (e.g., as is the case of the learner who appeared to decode the <gar> segment in <garmnent> by employing a small unit processing mechanism, but who appeared to decode the <ment> segment automatically, as shown by her comment that she read the <ment> segment 'altogether').

The results also support empirical findings in the L1 and L2 reading literature. For example, in her review of the L1 literature, Ehri (2015) reported four key decoding processes: sight word/automatic processing; decoding (which she used to refer to letter by letter/letter-cluster processing); analogizing; and prediction (which involves "using partial letters and context cues such as pictures or adjacent words in a sentence to anticipate the identity of an unfamiliar word (p.294)). Clearly, all of these processes were found in the present study (though labelled differently), except the prediction strategy. This is not surprising, however, as the items used in the study were presented individually rather than in a sentence context. Furthermore, within the L2 literature, and more specifically the qualitative literature, there is some similarity between the findings of the present study and those obtained by Woore (2010), as four out of the seven categories of the present study were reported in Woore (2010): namely, small-unit processing, large-unit processing, analogy-guided processes, and referencing L2 GPC knowledge (though Woore (2010) used slightly different labels). The other three processes, automatic processing, intuition-guided processing, and partial processing, were not reported in Woore (2010). However, it is worth noting that Woore (2010) did mention that some of his participants' comments appeared to reflect automatic processing but that comments of this nature were excluded from his analysis, hence the absence of the category. As for the intuition-guided and partial processing categories, it is not clear why these were observed in the present study but not in Woore's (2010). However, it is important to remember that the

[22] Though note that while approximately 80% of the comments under the automatic processing category were produced when reading real words, approximately 20% were produced when reading pseudowords, which the dual-route assumes would be read using the indirect route.

small sample size in both studies makes it difficult to draw generalisations. Another possible explanation may lie in the way in which the think-aloud task was implemented. For example, in the present study, comments underlying "partial processing" often emerged following further probing from the researcher (for example, 'And how about these letters?'), something which Woore (2010) may not have done. Additionally, two of the categories reported in Woore (2010) were not observed in the present study. These are: (1) subvocalizing possible pronunciations (i.e., where a participant attempts to produce intermediate vocalisations for the target items; and (2) making explicit comparisons with the L1 (English) decoding system. While it is not clear why "subvocalizing" was not reported in the present study, it is interesting that the present study participants did not make any explicit comparisons between the Arabic and English writing systems, potentially because the scripts embodying the two WS are different. This suggests that similarity in script, while facilitative in some respects, may sometimes be debilitative, as it may trigger L1-based GPC mappings (Woore, 2013).

Secondly, at a micro level, in-depth analysis of the comments produced in each processing category provided further insights into learners' decoding difficulties. To begin with, in line with findings from the analysis of participants' decoding outputs and experiences, learners' verbalisations clearly showed inadequate L2 GPC knowledge, which supports previous observations in the literature on Saudi ESL learners' decoding difficulties (e.g., Alsubaie, 2014). Furthermore, students seemed to employ a small-unit processing mechanism, when larger unit processing is required. It is not clear why this processing behaviour was present, and the lack of control groups makes it difficult to make claims about L1 transfer. However, it seems likely that it reflects learners' lack of awareness of large L2 spelling units, and the fact that they require holistic, rather than letter-by-letter, processing (Ehri, 2015). Learners also appeared to have inadequate knowledge of English syllabification rules, which was seen in their non-canonical segmentation of English words (for example, imagine a learner breaking <bone> into <b> and <one>, and then producing the realisation /bi:wʌn/) (this realisation was observed in the practice session of the decoding test). This suggests that knowledge of L2 GPC mappings may not be sufficient, as readers need to possess knowledge of L2 syllabification rules. Additionally, learners appeared to encounter difficulties making adequate analogies. This was often evident in: (a) their failure to identify a word (analog) that shares graphemic similarity with the target item (for example, when attempting to decode <budget>, a reader drew an analogy to <bought> and then assumed that <g> in <budget> would be silent), and/or (b) their attempts to draw analogies to "exception" words (for example, when attempting to

decode the pseudoword <pave>, a reader drew an analogy to <have>, an exception word, rather than to <save>, a regular word representing the most frequent grapheme-phoneme pattern). These difficulties that L2 decoders encounter with drawing adequate analogies have been previously reported in the literature (e.g., see Woore, 2010). Also, on occasion, there was a tendency to rely on "intuition" rather than reason when converting graphemes to phonemes (for example, as in the reader who omitted a phoneme because she "*felt it would suit the word more if I leave it"*. It is not clear what exactly drives these intuition-based judgements. However, it is possible that lack of secure GPC knowledge drives learners to rely on potentially unreliable sources of information "i.e., feelings or guesswork", as expressed by one student: "*there are words you feel that's how they should come. You can't think because you don't know much*". However, it should be noted here that on occasions these intuitions were successful. Thus, it is possible that, sometimes, they may reflect some form of still-developing GPC knowledge which learners are not consciously aware of (and thus unable to report on), as expressed by one learner:

- *S12: <knap> /knæp/ or maybe /næp/ because the k maybe silent because I came across a word but I don't remember what that word was and it was like that. The k was silent. So it could be /knæp/ or /næp/.*
  Researcher: which would you say is correct?
  S12: I don't know… I feel /nap/… but I also feel /knap/.

This learner seem to rely on intuition to help her decode <knap> ("*I feel /nap/… but I also feel /knap/")*, but that intuition appears to be based on insecure L2 GPC knowledge: "*because I came across a word but I don't remember what that word was and it was like that. The k was silent"*.

Thirdly, and although it was beyond the aim of the think-aloud task, it was interesting to see no indication of 'consonantal' processing behaviour. This observation tentatively challenges the "vowel blindness" claim put forward by Ryan and Meara (1991), and supports empirical evidence from Alhazmi, Milton, and Johnston's (2019) eye-tracking study, insights from the decoding outputs data presented earlier, and also the researcher's critical review of the conceptual validity of the vowel blindness hypothesis (see section 2.2.5.1.1.2).

6.1.3.2.Implications for theories of reading

Findings from the think-aloud task have some implications for reading theory. For example, and with respect to the dual-route framework, the processing categories reported appear to mirror the basic principles of the framework. Nonetheless, it may be useful to note

that the "indirect route" of the dual-route framework tells us *what* readers do when following this route (i.e., convert graphemes to phonemes) but lacks detail as to *how* they carry out this conversion process. The present data suggests that learners may do this by reliance on prior GPC knowledge, analogy making, and possibly by engaging in some level of guesswork (intuition-guided processing). Incorporating these observations in discussions of the dual-route framework could provide greater detail to our theoretical understanding of the decoding process.

Furthermore, certain patterns in the data appear to lend support to the main premise of connectionist reading models. To illustrate, one could argue that the thinking involved in some of the "assembled processing" categories may reflect weak connections between orthographic and phonological units. Similarly, the lack of thinking evident in the comments under the "automatic processing" category (for example, "I just said it") might reflect strong connections between orthographic and phonological units. However, one observation was found interesting in the present study and that is that the presence of strong connections between orthographic and phonological units (possibly evident in comments along the lines of "I just said it", "I did not have any thoughts, etc.) was not necessarily associated with the production of accurate decoding outputs (i.e., automatic but inaccurate decoding). Although such observations may not be unique to the L2 context, there are reasons to argue for their potential pervasiveness among L2 decoders (as opposed to L1 decoders), which in turn could inform discussions of connectionist models in the L2 reading context. To illustrate, it is worth remembering that connectionist models involve a learning procedure (called backpropagation) in which the decoding output produced by the model is matched against a correct output, with the model's performance improving following trial and error learning using a large corpus of grapheme-phoneme mappings (Seiderberg, 2007). The use of backpropagation, as Seidenberg (2007) noted, raises an important question regarding the validity of the model: "does human learning involve anything like comparing one's behavior to a target that fully specifies the correct response? (p.4)". Seidenberg (2007) proceeded to argue that backpropagation does take place in real life, as a child's decoding skill develops in response to two main teaching signals: (a) the teacher, as he puts it: "For tasks such as reading, for which there is explicit instruction usually, a target is often provided by a human teacher (p.5)"; and (b) self-generation of the target, where a child learns a grapheme-phoneme mapping following successful decoding of a novel written word, whose oral form has been previously acquired (often referred to as the self-teaching mechanism (Share, 1995).

While it is reasonable to acknowledge these teaching signals in the L1 reading context, it is less so in the L2 context (given presumed lack of phonics instruction as well as learners' potentially modest oral vocabulary size). What this means is that L2 readers may engage in unsuccessful decoding attempts, and given the lack of a "target" to match their pronunciation against, they may gradually end up self-learning inaccurate grapheme phoneme mappings. Consequently, the links connecting graphemes to phonemes in the mind of the L2 reader may be well entrenched, but not necessarily correct. This may explain why the "automatic but inaccurate" processing behaviour was observed in the present study. This observation highlights the need to consider these issues in SLA discussions of connectionist reading models.

Finally, and as noted previously, the lack of evidence for a consonantal processing behaviour has implications to Koda's (2013) Orthographic Distance framework. Because these have been mentioned in section 6.1.2, they are not reiterated here.

6.1.3.3.Implications for language pedagogy

Findings from the analysis of learners' decoding processes have several implications for pedagogy. First, the observation that automatic processing was not necessarily associated with the production of accurate decoding outputs suggests that, in the mind of the L2 decoder, connections between orthographic and phonological units may be very strong but not necessarily accurate. What caused the forming of these connections is unclear, but it is likely the outcome of inadequate instruction or self-learning. In either case, effort should be made to ensure language teachers, and particularly non-native users of the language, possess good L2 decoding knowledge, and to encourage L2 learners to verify the accuracy of their self-learned decoding realisations. Second, the fact that lack of adequate L2 GPC knowledge was also observed in learners' verbalisations further stresses the need to implement phonics instruction in L2 settings. Third, it may be useful to familiarize students with frequent large spelling units/morphemes (for example, <-tion>, <-ment>) and to encourage students to process these holistically rather than letter by letter. Fourth, as students tended to segment words incorrectly, potentially explaining some of the decoding inaccuracies, students may benefit from learning some L2 syllabification rules (for example, that a syllable includes one vowel sound, that a bound morpheme (e.g., de-) is a syllable in its own, etc.). Teachers could also ask students to divide words into syllables and then blend the divided syllables together to decode words. Furthermore, since readers appear to engage in analogy-making to help them decode words, teachers could teach students how to make analogies effectively, to ensure successful application of this strategy. Additionally, teachers may want to raise students' awareness of the

need to use reason rather than guesswork or intuition when decoding. Finally, teachers may also inform students of the need to process the whole word (i.e., all the letters contained within a word), rather than only parts of it, when attempting to decode words.

6.1.3.4.Limitations of the findings

The think-aloud findings provided rich insights into the decoding processes and difficulties of L2 readers. However, using think-aloud protocols as a mean to investigate cognitive processes is not without problems. As noted earlier, comments along the lines of "I just said it, I know it", which constituted around a quarter of the total comments observed, were taken to represent automatic processing. Although patterns in the data clearly support this interpretation (for example, most of the responses in this category were produced when encountering real words), it is difficult to fully ascertain whether the comments produced were the outcome of genuine automatic processing or difficulties verbalising thoughts, the latter being a recognized limitation of think-alouds. In fact, approximately 20% of the comments in this category were produced when reading pseudowords, which are less likely to be processed automatically. However, it is important to remember that automaticity should be conceptualized as a continuum rather than a dichotomous entity (Logan, 1997) and it is highly likely that learners' decoding of these pseudowords was too quick to be reported on. This is further supported by the observation that almost all of the pseudowords processed "automatically" were short in length (for example, <ip>, <om>, <lat>, etc.).

Furthermore, although the task instructions explicitly asked students to verbalise their thoughts retrospectively, on occasions, students verbalised their thoughts concurrently (i.e., as they were attempting to decode the item). One issue with concurrent verbalisations is that they may interfere with normal task (decoding) performance, which in turn may impact the validity of the processes reported.

Notwithstanding these limitations, it is important to mention that there was tentative evidence for the validity and reliability of the method. For example, there was often, though not always, a clear match between the verbalized thoughts and the decoding output generated, which can be taken as evidence for the predictive validity of the method. Furthermore, the mere fact that the processes reported largely corresponded with reading theory may be seen as supportive of the validity of the method. Additionally, there was a high degree of consistency in the processes reported across the participants, and across different samples of learners in other studies (for example, English learners of French in Woore, 2010), which can be taken as evidence for the reliability of the method.

Now that the findings relating to the decoding abilities, realisations, experiences, and processes have been discussed, the next section discusses the findings relating to the contribution of second language decoding proficiency to second language reading comprehension.

6.2. The contribution of second language decoding proficiency to second language reading comprehension

6.2.1. The findings on the contribution of L2 decoding proficiency to L2 reading comprehension and the literature

6.2.1.1. The Contribution of L2 decoding proficiency to L2 reading comprehension

The findings obtained were supportive of those found in the L2 literature, in the sense that decoding was found to be a statistically significant predictor of reading comprehension (Koda, 1998; Jeon, 2011; Jiang et al 1998; and Jiang, 2016). However, there is considerable variability with respect to the size of that contribution across different studies. Overall decoding proficiency explained only about 1% of the variability in reading comprehension in the current study, whereas its contribution in previous studies has ranged from 11% (Jiang et al., 2012) to 65% (Jiang, 2016). This variability across studies is also present in the L1 reading context. For example, in García and Cain's (2014) review, the authors reported that in some studies (e.g., Berninger, Abbott, Vermeulen, & Fulton, 2006) the contribution of decoding was negligible (R^2= less than 1%), while in others (e.g., Katzir, et al., 2006), decoding ability almost fully predicted L1 reading comprehension ability (R^2= 90%).

There might be several reasons for this discrepancy, including the analytical method applied, the L1WS of the participants, other contextual variables (e.g., country of residence), characteristics of the reading comprehension measure, the proficiency level of the learner, and the decoding measure used. These are discussed in detail below with reference to findings from the present study.

- **The analytical method**

It is very important to remember here that previous studies did not control for key predictors of L2 reading comprehension skill (for example, L2 vocabulary and grammar knowledge) when assessing the contribution of L2 decoding proficiency to L2 reading comprehension. As these skills tend to correlate with decoding skill, failure to control for these statistically may have overestimated the contribution of L2 decoding proficiency to L2 reading comprehension. In other words, the contribution of L2 decoding proficiency observed in previous studies may simply, or additionally, reflect variability in L2 vocabulary knowledge,

L2 grammar knowledge, etc. This is further supported by the finding that the contribution of decoding in the present study decreased from approximately 35% (when other variables were not controlled for) to approximately 1% (after controlling for other variables). If one was to consider the contribution of decoding before controlling for these other variables, then the percentage of variance found in the present study (35%) would be somewhat similar to that of Jiang's (2016) sample of Arabic L2 learners (46%).

- **The L1WS of the participants**

Another variable that may contribute to the variability in the findings obtained relates to the L1WS of the participants. It has been argued that the contribution of decoding would be larger where the L1 and L2 writing systems are similar than where they are different (for example, alphabetic L1WS and alphabetic L2WS). For example, Jiang (2016) tested Chinese (logographic L1), Japanese (Syllabic L1), Arabic (consonantal L1) and Spanish (alphabetic L1) English learners in a single study, and found that the decoding contribution was largest among the Spanish group (65%), followed by Arabic group (46%), the Japanese group (32%), and finally the Chinese group (18%).

Jiang (2016) did not explain why the contribution of L2 decoding proficiency to L2 reading comprehension was smaller among L1 Arabic readers, compared to readers with alphabetic L1WS backgrounds, though they did provide an overall explanation for their findings: the closer the L1WS to the L2WS, the stronger the relation between L2 decoding proficiency and L2 reading comprehension. In the case of Arabic learners of English, a possible explanation may relate to the role of the sentential context in the Arabic decoding process. As mentioned earlier, short vowels are absent from typical Arabic texts, and to successfully decode words, readers may rely on the sentential context, among other sources of information, e.g., lexical knowledge, to infer the missing phonological information (Frost, 2006) . Saigh and Schmitt (2012) mention that "given that Arabs' L1 literacy skills involves relying on syntactic, semantic and discourse contexts when recognizing words in texts, Arab ESL learners may overcome word recognition problems … by relying on context" (p. 30). If this is the case, it follows then that the presence of the sentential context in comprehension tests may boost the decoding performance of these learners in a way that traditional decontextualized decoding tasks cannot capture, which in turn may explain why traditional decoding measures contribute less strongly to reading comprehension among this group.

- **Study context: L1 vs. L2 speaking country**

Another reason for the discrepancy in the findings may relate to the study context, and particularly, whether the L2 readers were based in in the first or second language (English) speaking country. To illustrate, it is possible that L2 readers residing in the L2 speaking country would have greater auditory vocabulary size than those based in their home countries. Since one benefit of improving decoding proficiency is to facilitate access to oral vocabulary, variability in decoding skill among readers based in the L2 country may have a greater effect on their reading comprehension, relative to readers based in their home countries.

- **Reading comprehension measure**

Another variable that has been argued to affect the relationship between decoding and reading comprehension is the method of reading comprehension assessment, and particularly, whether or not it imposes a time limit (García and Cain, 2014). Particularly, the contribution of decoding to reading comprehension is assumed to be weaker where the reading comprehension measure is untimed, as less skilled decoders will have the opportunity to re-read the text, correct their decoding, which could in turn improve their reading comprehension. In the present study, the reading comprehension measure was timed; however, that time limit was potentially sufficient to allow for any 'revisions', as most participants tended to complete the test before the time limit was reached. Unfortunately, previous studies did not provide enough information as to whether or not their reading comprehension tests imposed a time limit, but if they did, then this might also explain the relatively small contribution found in the present study (though note that the contribution was small (i.e., 1%) after controlling for key predictors of reading comprehension which previous studies did not consider).

Aside from variability in the analytical method employed, the L1WS background, and study context, two variables, the L2 proficiency level of the learner and the decoding measure used, could also play a role in moderating the size of the decoding contribution. As the present study explicitly investigated these two variables, they will be discussed in the next two sections.

6.2.1.2.The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the proficiency level of the learner

As expected, results from the present study suggested that the contribution of L2 decoding proficiency varies across readers at different proficiency levels. Particularly, the contribution of L2 decoding proficiency appeared to decrease as linguistic proficiency increased, a finding that is in line with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). A possible explanation for this finding may relate to what is known as "compensatory processing" (Stanovich, 1980; Stanovich, 1986; Stanovich, 2000). To illustrate, given their

more sophisticated linguistic knowledge, readers with higher levels of proficiency (i.e., vocabulary and grammar knowledge) might be less disrupted by poor decoding than learners with lower levels of linguistic proficiency. This is because high proficiency learners are more likely to be able to compensate for any decoding deficiencies (for example, by being more skilled at inferring word meaning from the sentential context, thus being less disrupted by poor decoding, and/or more able verify the accuracy of their decoding realisations).

This work was the first to examine the role of proficiency in moderating the contribution of L2 decoding proficiency to L2 reading comprehension, and as such, it is not possible to see how the findings obtained relate to existing L2 literature. However, findings from previous individual studies appear to suggest a similar trend. For example, Koda's (1998) study, which reported a large decoding contribution (50%), included beginner-level learners, whereas Jiang's et al. (2012) study, which reported the smallest decoding contribution, recruited intermediate-level learners (though it remains difficult to make strong conclusions from individual studies, as several other variables, other than the proficiency level, may play a role). Furthermore, it is interesting to mention here a relevant finding from the L1 reading context. In a meta-analysis by García and Cain (2014), the listening comprehension level of the reader, which can be a proxy for overall linguistic proficiency, moderated the size of the decoding and reading comprehension relationship. Particularly, they found that the contribution of decoding was smaller in learners with high levels of listening comprehension, and larger in learners with low levels of listening comprehension, a pattern that resembled findings from the present study.

Finally, it is worth noting here that learners' proficiency level appeared to moderate the contribution of decoding regardless of the decoding measure used, i.e., a composite measure of real word and pseudoword decoding (overall L2 decoding proficiency), a real word measure only, or a pseudoword measure only (though in the case of the pseudoword measure, that pattern was approaching statistical significance, p=.06).

6.2.1.3.The contribution of L2 decoding proficiency to L2 reading comprehension as a function of the L2 decoding measure

The two key findings from this part of the study were that (1) the contribution of decoding to reading comprehension appeared to vary as a function of the decoding measure (though see below); and that (2) contrary to expectations, the two decoding measures did not make a unique contribution to L2 reading comprehension performance.

That the contribution of L2 decoding proficiency to L2 reading comprehension varied as a function of the decoding measure corroborates the findings of previous work in the L1 context.

For example, Nation and Snowling (1997) found that their two decoding measures (a pseudoword decoding measure, and a real word decoding measure) contributed differently to reading comprehension, with the former making a smaller contribution to reading comprehension, compared to the latter. Similar findings were also reported in Florit and Kain (2011) and García and Cain's (2014) meta-analysis and the pattern of the results was also the same. That is, the relationship between measures of real word decoding ability and reading comprehension were stronger than the relationship between pseudoword decoding ability and reading comprehension. This result may be explained by the fact that semantic knowledge can influence performance on the real word decoding measure but not on the pseudoword measure (Nation and Snowling, 1997). Imagine two learners (A and B) attempting to decode the word <atmosphere> for the first time. Learner A understands what <atmosphere> means from prior exposure to its auditory form, whereas learner B does not. When encountering the written form of <atmosphere>, this prior semantic knowledge might inform (and potentially boost) the decoding performance of learner A. This is further supported by the finding that when the two decoding measures were entered after vocabulary knowledge, their contributions to treading comprehension was very similar (ΔR=.016 for the pseudoword measure and ΔR=.018 for the real word measure). (before accounting for vocabulary knowledge, and other reading components, the contributions of the two measures were different, 42% and 21%, for the L2 real word decoding and L2 pseudoword decoding measures, respectively).

The finding that the two measures did not contribute uniquely to reading comprehension was interesting. This result is contrary to that reported by Jiang, Sawaki, and Sabatini (2012), where real word decoding explained an additional variance in reading comprehension after controlling for pseudoword decoding (note that Jiang et al. did not switch the order of entry, i.e., did not enter pseudoword decoding after real word decoding). However, it is important to note three issues when discussing Jiang et al's (2012) results: (1) pseudoword decoding lost its effect when real word decoding was entered in the model, which suggests that only real word decoding, but not pseudoword decoding, makes a unique contribution to L2 reading comprehension (though note that the *p* value of the real word measure, when entered after pseudoword decoding, was almost nonsignificant (p=.04)); and most importantly, (2) Jiang et al (2012) did not include any control variables. It is very possible that if vocabulary knowledge had been assessed and included in the model – along with pseudoword decoding – real word decoding proficiency would have lost its modest unique effect. In fact, in the present study, real word decoding proficiency and pseudoword decoding did make unique contributions to L2 reading comprehension but that was *only* when other variables were not controlled for.

Two reasons may explain the lack of an independent contribution. First, real word and pseudoword decoding skills tend to develop more or less in tandem (Joshi (2000) argues that sight word decoding is potentially built on pseudoword decoding skill, and that, therefore, the two might not be independent of one another). Statistically, this relatively concurrent development would be manifested in a large correlation between the two measures, making it difficult to assess their contributions independently of one another. Second, the real word measure in the present study did not strictly include *familiar* items. Therefore, it is possible that students engaged in some level of phonological assembly while reading some of the real word items. However, it may be worth mentioning here that following these findings, the researcher decided to subject TOWRE's real word decoding test to word frequency analysis using Lextutor (Cobb, 2019) (see Appendix F), and to see the frequency bands of the items read by the participants in the present study. The analysis showed that 94% of the items that students read belonged to the most frequent 1000-word family. As Saudi foundation year ESL students would have learned English for a minimum of 8 years in schools, there was no reason to question their prior encounter with these highly frequent words. Thus, and while it is difficult to completely eliminate the possibility of measurement oversight, it is highly likely that the absence of an independent contribution is the outcome of the two skills being intrinsically related.

6.2.1.4.The contribution of L2 decoding proficiency to L2 reading comprehension: insights from Dominance analysis and Relative weight analysis

With respect to the relative importance of overall decoding proficiency, results from Relative Weight and Dominance Weight analyses appeared to replicate findings from the traditional regression analysis, as overall decoding proficiency continued to emerge as the third strongest predictor for L2 reading comprehension, following L2 grammar knowledge and L2 vocabulary knowledge (and preceding L1 reading comprehension ability). These findings are also similar to findings from Jeon and Yamashita's (2014) meta-analysis, which suggest that (1) grammar and vocabulary knowledge are stronger predictors of L2 reading comprehension, relative to decoding proficiency; and (2) that decoding proficiency is a stronger predictor of L2 reading comprehension, relative to L1 reading comprehension ability.

However, when real word decoding proficiency and pseudoword decoding proficiency were included in one model (along with the other predictors) using traditional regression analyses, neither emerged as a significant predictor of L2 reading comprehension. This is despite the medium to large correlation that they shared with L2 reading comprehension,

suggesting the possibility of a collinearity issue. The researcher's concerns were reinforced by another finding. Whenever one of the two variables (real word decoding or pseudoword decoding) was removed from the model, the other would emerge as a significant predictor of L2 reading comprehension. Thus, if only traditional regression analysis were used, we might have overlooked the importance of decoding proficiency as both decoding measures would emerge as nonsignificant predictors of reading comprehension when included in one model. This observation is similar to that of previous L1 reading studies, where traditional regression analyses appeared to underestimate the importance of certain reading components (e.g., see Mellard et al, 2012). Furthermore, in terms of relative importance, traditional regression analyses (and the resulting beta weights) placed real word decoding proficiency and pseudoword decoding proficiency as the weakest predictors of reading comprehension (i.e., they were not only preceded by grammar knowledge and vocabulary knowledge, but also L1 reading comprehension ability). The dominance weight and relative weight analyses, however, which were better suited to deal with collinear data, generated slightly different and arguably more meaningful findings. Real word decoding proficiency and pseudoword decoding proficiency were ranked as more important for L2 reading comprehension than L1 reading comprehension ability. These findings replicate those found in the analytical literature, which questioned the reliability of beta weights in the context of correlated predictors (Kraha, Turner, Nimon, Zientek, & Henson, 2012).

It is also interesting to note that the difference between the relative weights of vocabulary and grammar knowledge was not statistically significant. This finding is similar to that reported in Jeon and Yamashita's (2014) meta-analysis, where the confidence intervals of L2 grammar knowledge and L2 vocabulary knowledge overlapped significantly, leading the authors to conclude that both may be equally important for L2 reading comprehension.

6.2.1.5. The contribution of L2 decoding proficiency to L2 reading comprehension performance: qualitative insights

The interview findings suggested that L2 decoding proficiency may interfere with learners' reading comprehension. This finding replicates observations from previous primarily qualitative studies (e.g., Erler, 2003; Alsubaie, 2014), and provides additional support for the quantitative findings obtained in the present study.

Furthermore, in-depth analysis of learners' comments provided first-hand explanations for the quantitative findings obtained. Poor decoding skill appeared to disrupt meaning access, modulate situational affect, and possibly overload cognitive capacity. (A note of caution is due

here, however, as comments relating to overloading cognitive capacity were reported following further probing from the researcher, rather than spontaneously). These accounts – specifically those relating to facilitating meaning access and freeing cognitive capacity – mirror the theoretical literature into the mechanisms underpinning the role of decoding proficiency in reading comprehension (see section 2.1.2.2.6.1 in the literature review chapter). However, the theoretical literature also discusses the role of accurate phonological codes in enhancing working memory functioning (Perfetti, Zhang, & Berent, 1992), a mechanism that was not observed in learners' interviews, possibly due to learners' lack of awareness of this process. However, it is relevant to mention here that students reported a tendency to ignore difficult-to-decode words. As phonological codes are useful for recreating prosodic aspects (for example, sentence boundaries) (Pollatsek, 2015), such reading behaviour might disturb building prosodic representations, which in turn might interfere with their comprehension. Additionally, it is interesting to note that the interview findings suggest that an "affective mechanism" may also be at play. Current theoretical accounts into the mechanisms underlying the contribution of decoding were mostly cognitive in nature, with the affective dimension being largely overlooked.

Finally, the regression model tested in the present study was able to explain around 63% of the variability in L2 reading comprehension performance, leaving around 37% of the variance unexplained. The qualitative findings were able to provide insights into what might explain this remaining variance, as factors other than those tested emerged in learners' interviews. These included affective state, background/world knowledge, morphological knowledge, text length, time pressure, and concentration issues. Some of these factors may overlap, of course. For example, lengthy texts may be problematic because they require a high attention span, which students with concentration difficulties may find problematic. Lengthy texts may also be problematic because a certain level of motivation may be needed to support task completion. Concentration issues may also arise in the presence of a negative affect (e.g., anxiety) or may simply reflect working memory deficits. Though it is beyond the scope of this thesis to discuss these variables in depth, it is useful to mention that most of these were reported in previous studies (e.g., see Zhao, 2013; Guthrie et al., 2007, Jeon, 2011; Daneman and Merickle, 1996, and Langer, 1984, respectively, for evidence for the role of anxiety, motivation, morphological knowledge, working memory, and background knowledge in reading comprehension).

6.2.2. Implications for theories of reading

The findings of the third research question have several implications for theories of reading. First, they have the potential to contribute to a componential model of L2 reading comprehension, as findings suggest that a componential model of L2 reading comprehension should take into account decoding ability (note that Bernhardt's (2000; 2005) widely-cited model, includes the components of grammar knowledge, vocabulary knowledge, and L1 reading comprehension ability, but makes no mention of decoding ability). Furthermore, a componential model of L2 reading comprehension should take into consideration that the contribution of L2 decoding proficiency is not stable across L2 development, but rather decreases as L2 learners progress to higher levels of L2 proficiency. Additionally, componential models of L2 reading comprehension should take into account the fact that componential skills do not contribute equally to reading comprehension, as some components appear to be more crucial for successful comprehension than others. For example, L2 grammar and vocabulary knowledge appear to be more predictive of L2 reading comprehension performance, relative to L2 decoding proficiency and L1 reading comprehension ability.

The findings obtained also have the potential to contribute to our understanding of the mechanisms that underpin the contribution of L2 decoding proficiency to L2 reading comprehension. Current theoretical understanding suggests that three mechanisms underpin this contribution; the first is by means of facilitating meaning access; the second relates to the role of effortless decoding in freeing cognitive capacity; and the third relates to the role of phonological codes in enhancing working memory functioning, all three reflective of a purely cognitive perspective on reading comprehension. However, the present study tentatively suggests that an affective mechanism may also be at play. That is, decoding deficiencies may modulate situational affect, which in turn may interfere with reading comprehension performance.

Finally, findings from interviews with learners suggest that variables other than grammar knowledge, vocabulary knowledge, L2 decoding proficiency, and L1 reading comprehension ability could potentially contribute to L2 reading comprehension. These include: affective state (for example, motivation and anxiety); time constraints; background/topic knowledge; morphological knowledge; and inferencing ability. Thus, consideration of these components might also be needed in componential models of L2 reading comprehension (Also see Martínez, 2009; Saito, Horwitz, & Garza, 1999; Zhang & Koda, 2012).

6.2.3. Implications for language pedagogy

The present study does not permit identification of causal relationships. However, if one is to postulate a dependence of L2 reading comprehension on L2 decoding proficiency, an implication for pedagogy would be that focused efforts to improve L2 decoding knowledge may result in relatively more efficient reading comprehension. Furthermore, for the purposes of improving reading comprehension, decoding instruction should focus on improving learners' decoding accuracy as well as speed. As discussed in detail in previous sections, it is possible that decoding accuracy could be enhanced by introducing phonics instruction (see sections 6.1.1.3. and 6.1.2.3 for relevant pedagogical implications). Decoding speed, on the other hand, could be improved by implementing a range of decoding speed interventions, including assisted reading (for example, where learners read silently while the language teacher reads words outloud), repeated reading (for example, where a learner reads a list of words repeatedly until they can reach an appropriate reading time criterion), and recorded reading (for example, where a learner plays an audio recorded list of high frequency words, and reads along with the recording, before they proceed to read the list on their own) (Grabe, 2009; Rasisinsky, 2003)). In fact, there is evidence for the effectiveness of such interventions on improving L2 readers' decoding speed (Tan, Moore, Dixon, & Nicholson, 1994).

Beyond improving decoding accuracy and speed, language learners need to be made aware of the need to decode all words in a text, even if these are perceived to be difficult. Building a habit of bypassing difficult-to-decode words can be problematic, not only because it impedes opportunities for self-teaching (Share, 1995), but also because such behavior can interfere with prosodic building (i.e., the building of mental representations of phrase and sentence boundaries) thus potentially disturbing reading comprehension (Pollatsek, 2015). Additionally, the findings suggest that the reading comprehension of beginner and elementary level learners may benefit more from this decoding instruction, as more proficient learners may potentially be able to compensate for their decoding difficulties.

Nonetheless, the finding that L2 vocabulary knowledge and L2 grammar knowledge were stronger predictors of L2 reading comprehension than was L2 decoding proficiency suggests that more time and effort should be devoted to improving learners' L2 vocabulary and grammar knowledge, if the goal is to improve their L2 reading comprehension. Finally, having underscored the major roles of vocabulary knowledge, grammar knowledge, and decoding ability, one must not overlook the roles of other cognitive (for example, working memory), linguistic (for example, morphological knowledge), and affective factors (for example,

motivation and anxiety), as these could collectively contribute to better reading comprehension outcomes.

6.2.4. Limitations of the findings

As with any research study, the present one has its weaknesses and limitations. The limitations discussed here concern the characteristics of the sample tested, the data collection methods, and the analytical techniques performed. Each of these will be discussed in turn below.

- Characteristics of the sample

With regard to the sample, due to time and practicality constraints, it was not possible to collect data from more than one higher education setting. Thus, findings of the whole study (i.e., not just part 2) may not be readily generalisable to low level English learners in other Saudi Arabian universities. Nonetheless, and as mentioned previously, there is no reason to assume that these learners would be very different from students in other institutions. Furthermore, due to learners' other commitments and packed schedules, it was not possible to achieve random sampling. Further, the present study included low level English learners only and it is open to question if similar findings would be obtained from learners in the higher half of the proficiency scale.

- The study measures

Limitations relating to the measures used in the present study are also worth noting. Due to time and practical constraints, most of the variables included in the present study were assessed using single measures. It is possible that using multiple measures for each skill would enhance skill assessment, while also allowing for the construction of latent variables, which in turn would facilitate the use of more advanced statistical techniques, for example, structural equation modelling. Furthermore, the decoding measures used in the present study did not allow for individual assessment of the contributions of decoding accuracy and speed. It would be interesting to see if the two (accuracy and speed) would individually contribute to variability in L2 reading comprehension performance, as well as to see their respective importance relative to each other. For example, in the L1 context, García and Cain (2014) found that measures of decoding accuracy correlated more strongly with reading comprehension than measures of decoding speed, which potentially suggests that the impact of decoding deficiencies on achieving semantic access may be greater than their impact on overloading the limited cognitive capacity system. Another limitation also relates to the real word decoding measure used. The measure used in the present study did not strictly include familiar real words. It may

be useful to see if a real word measure strictly including highly familiar items would contribute to variability in reading comprehension. However, and as noted previously, one concern in using a test of highly familiar words is that it could lead to a ceiling effect, and this lack of variation in scores would consequently lead to lack of variability in reading comprehension scores (a problem often reported in L1 reading research). Another limitation may also relate to the fact that the decoding measure was not contextualized. It is possible that the performance of Arab ESL readers on single word/pseudoword tasks does not exactly represent their decoding performance when reading texts, as the latter would allow for the use of the sentential context to aid decoding. However, it should be noted that one issue in using contextualised tests is that they may conflate decoding ability with other language abilities, such as vocabulary and grammar knowledge, making it difficult to rigorously assess the contribution of decoding proficiency. This is not to mention that the use of contextualised tests would reduce the ability to compare findings as most previous studies tended to use decontextualised decoding tests.

- Design and analytical limitations

The present study included key predictors of L2 reading comprehension (L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability) and implemented multivariate regression techniques to reduce error variance when assessing the contribution of L2 decoding to L2 reading comprehension. However, the correlational nature of the design still makes it difficult to make causal claims about the impact of L2 decoding ability on L2 reading comprehension.

Furthermore, the present study was limited in that it was mainly concerned with the contribution of L2 decoding proficiency to L2 reading comprehension. However, the relationship between L2 decoding proficiency and L2 reading comprehension is likely to be reciprocal. That is, L2 reading comprehension may also contribute to variability in L2 decoding performance, as better comprehenders are likely to engage in extensive reading, which ultimately could improve their decoding abilities (the so-called 'Matthew effect': see Stanovich 1986).

7. Conclusion

The present study investigated Saudi foundation year ESL students' (CEFR: A1, A2, B1) decoding abilities, realisations, experiences, and processes, as well as the contribution of their second language decoding proficiency to their second language reading comprehension performance. This is the first study to examine the L2 decoding abilities, realisations, experiences, and processes of Saudi foundation year ESL students. It is also the first study that assessed the contribution of L2 decoding proficiency to L2 reading comprehension after controlling for key predictors of L2 reading comprehension performance, namely, L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability.

In this work, a total of 212 students took part in a real word and pseudoword decoding task which aimed to assess their decoding proficiency. The decoding realisations of a subsample of participants (N=60) were then phonemically transcribed and subjected to error analysis. A complementary qualitative investigation was also carried out, in which a subsample of learners took part in an interview which aimed to understand their decoding experiences (N=12) as well as a think-aloud task which explored their decoding processes (N=15). To rigorously assess the contribution of their L2 decoding proficiency to their L2 reading comprehension performance, additional measures of L2 vocabulary knowledge, L2 grammar knowledge, L1 reading comprehension ability, and L2 reading comprehension were administered to the same 212 participants. A complementary qualitative investigation was also carried out in which 18 participants took part in interviews aimed at validating and providing further insights to the quantitative findings obtained.

Analysis of learners' performance on the decoding tests suggested that they encounter difficulties with decoding the second language, and that, not unexpectedly, lower proficiency learners encounter more decoding difficulties than higher proficiency learners. Learners also encountered more difficulties reading pseudowords than real words. A thorough error analysis of their decoding realisations highlighted difficulties with decoding vowel graphemes and consonant graphemes, a tendency to add to, omit, and mis-order words' phonemes, along with a tendency to make whole-word substitutions. Furthermore, in depth analysis of the nature of these realisations provided interesting insights. For example, with respect to the vowel graphemes, error analysis of participants' realisations showed a clear lack of "randomness" in decoding these vowels, and further suggests that vowel decoding errors may have resulted from deficient knowledge of L2 vowel GPC and/or vowel articulatory difficulties. With respect to consonants, the results seemed to suggest that consonant decoding difficulties were potentially

the result of articulatory difficulties, lack of familiarity with the L2 script, and lack of L2 GPC knowledge. Where it was possible to identify a pattern in the phoneme insertion, addition, and misordering cases, the results suggest that these were potentially the result of lack of L2 GPC knowledge, and/or articulatory difficulties. Furthermore, the present work showed that, at a macro level, no qualitative differences appear to exist in the nature of decoding difficulties between participants at different proficiency levels (A1, A2, B1), as the categories of error identified were observed in all three groups of proficiency. However, quantitative differences were found between the groups: the error rate in each of the categories seemed to decrease as proficiency level increased. Furthermore, analysis of learners' think-aloud protocols yielded several interesting insights. Learners seemed to engage in a range of decoding processes. These included: (a) automatic processing; (b) assembled processing – evident in small-unit processing, large-unit processing, conscious computing of words' phonology by referencing L2GPC knowledge, making analogies (analogy-guided processing), and/or drawing on intuitions (intuition-guided processing) – and (c) partial/incomplete processing. In-depth analysis of the comments observed in each of these processing categories validated findings from learners' decoding realisations and experiences, and provided additional insights into their decoding difficulties. Learners' verbalisations clearly showed deficient L2 GPC knowledge, a tendency to employ a small-unit processing mechanism when larger unit processing is required, minimal knowledge of English syllabification rules, a tendency to make unsuccessful analogies, and, on occasions, a tendency to rely on "intuition" rather than reason when converting graphemes to phonemes. In addition, it was interesting to find no clear evidence for a consonantal processing behaviour (i.e., processing consonants and ignoring vowels), which is in line with the insights drawn from the decoding realisations data.

The second part of the study aimed to assess the contribution of second language decoding proficiency to second language reading comprehension among this sample of Saudi foundation year ESL students. A series of regression analyses showed that L2 decoding proficiency correlated moderately with L2 reading comprehension performance, and that the contribution of L2 decoding proficiency to L2 reading comprehension was significant, even after controlling for the effects of L2 vocabulary knowledge, L2 grammar knowledge, and L1 reading comprehension ability. Additionally, the findings suggested that this contribution varied as a function of the proficiency level of the learner, as it seemed to decrease as learners' proficiency increased. Furthermore, the contribution of decoding to reading comprehension appeared to vary as a function of the decoding measure used but that is only when that

contribution was assessed before controlling for other key predictors of L2 reading comprehension. When key predictors of L2 reading comprehension were controlled for, the contributions of the two decoding measures were very similar in magnitude. Additionally, the two decoding measures did not significantly contribute to reading comprehension independently of one another. Finally, in complementary interviews, decoding proficiency emerged as one of the factors that interfere with learners' reading comprehension, which provided additional support for the quantitative findings obtained in the present study. Furthermore, in-depth analysis of learners' comments provided additional insights into the mechanisms that underpin the contribution of L2 decoding proficiency to L2 reading comprehension (poor decoding skill appeared to interfere with meaning access, modulate situational affect, and overload cognitive capacity) and into what might explain the remaining variance in the regression model tested (including affective state, background/world knowledge, morphological knowledge, text length, time pressure, and concentration issues).

All in all, and as previously discussed, these findings had several implications for theories of reading, including theories of L2 decoding (Koda's Orthographic Distance hypothesis), dual-route models of reading, application of connectionist models to the L2 context, and componential models of L2 reading comprehension (Bernhardt's model of L2 reading comprehension). The findings also had several implications for second language pedagogy, as they, among other recommendations, highlight the need to introduce decoding instruction in L2 classrooms and suggest areas that L2 decoding instruction should focus on in the Saudi ELT context.

However, the present study was not without limitations, many of which have been covered in the preceding chapter. These limitations provide insights, and serve as starting points, for future research. In the final section below, the researcher offers some recommendations based on, though not necessarily limited to, the limitations of the present study.

7.1.Recommendations for future research

7.1.1. Documenting the decoding abilities of second language learners

Previous research into L2 decoding was mainly concerned with the *nature* of the L2 decoding process (for example, how it differs across learners from different L1WS backgrounds, how the nature of the L1WS can accelerate/impede L2 decoding acquisition, etc.) (Nassaji, 2014). This left several essential issues about L2 decoding acquisition unclear. One major issue is the lack of understanding into the decoding abilities of learners throughout the

course of L2 development. The present study examined the decoding abilities of A1, A2, and B1 level learners, utilizing a widely recognised decoding assessment tool (TOWRE). Future research could quantify and document the decoding abilities of L2 learners at all developmental stages (A1, A2, B1, B2, C1, and C2), across different decoding assessment tools (for example, TOWRE and Woodcock), L1WS backgrounds (for example, logographic L1WS and alphabetic L1WS), age groups (for example, children and adult learners), and teaching contexts (for example, contexts where phonics instruction is/is not implemented), ideally using longitudinal designs. This research would yield fruitful insights into the theory of L2 decoding. For example, in the L1 context, decoding is considered a "constrained" skill, in that it can be mastered entirely and to a ceiling level (unlike vocabulary knowledge, for example, which can continue to grow, and hence is considered an "unconstrained" skill) (Paris, 2005). Investigating the decoding abilities of L2 learners at different proficiency levels would allow us to understand whether mastery of decoding skill is equally possible in the L2 context, and if not, then how far in the decoding mastery spectrum L2 learners can go, as well as whether this varies as a function of the variables mentioned above (i.e., L1WS of the learners, age group, teaching context, decoding assessment tool). This research would also be very useful for L2 decoding pedagogy, as it would help identify decoding developmental "milestones" that students' decoding performance could be matched against, leading to a more informed decoding pedagogy.

7.1.2. The nature of second language of decoding development

To date, much ambiguity surrounds the very nature of L2 decoding development. Previous studies in the L2 decoding literature were often small-scale, comprising learners from one proficiency level. As noted previously, this makes it difficult to see whether learners at different developmental stages would exhibit decoding difficulties of a similar/different nature. Future research could address these issues by examining the nature of L2 decoding realisations (also across the full span of L2 development (i.e., CEFR: A1, A2, B1, B2, C1, and C2)), ideally longitudinally. Such investigation would provide answers to several unanswered questions: for example, what facets of L2 decoding knowledge do learners acquire at the beginning stages of L2 development? What facets of L2 decoding knowledge do learners acquire at later stages of L2 development? Are there any facets of L2 decoding knowledge that L2 learners would not be able to acquire? If yes, what are these? Answering these questions would significantly enhance our understanding of the nature of L2 decoding development. Subsequent research

could also test any emerging frameworks across learners from different L1WS backgrounds and age groups.

7.1.3. Validity and reliability studies into the use of think-aloud protocols as means for investigating word decoding processes

Past L1 and L2 reading studies have used numerous techniques (for example, reaction times, error rate, and eye fixations) to study the cognitive processes involved when reading words. Despite the variety, these methods share a key characteristic: predictions about word reading processes are made *a priori*, with the tasks primarily designed to refute or support these pre-existing theoretical suppositions. Whilst this approach, largely quantitative and designed to test existing hypotheses, has undoubtedly advanced our understanding of the reading process, it might have missed important elements of that process. The present study used think-aloud protocols when investigating the decoding processes of L2 decoders and the methodology appeared to yield useful insights. Although much research has been done to investigate the validity and reliability of think aloud protocols in general (Ericsson & Simon, 1993), to the best of the researcher's knowledge, no research has been done to assess their validity and reliability as means for investigating decoding processes. Therefore, future research might explore various aspects of the validity and reliability of the method. For example, it would be useful to investigate the method's predictive validity (for example, by examining whether learners' think-alouds match their decoding performance), concurrent validity (e.g, by examining decoding processes using think-aloud protocols vs. eye tracking methods, and see whether similar results are obtained across the two methods), and whether any validity and reliability evidence varies as a function of the administration method (i.e., concurrent think alouds vs. retrospective think alouds).

7.1.4. Structural equation modelling studies into the contribution of second language decoding proficiency to second language reading comprehension

Future studies could use more advanced analytical techniques, such as structural equation modelling, when assessing the contribution of second language decoding proficiency to second language reading comprehension. Structural equation modelling has some advantages over multiple regression techniques. It allows for simultaneous estimation of complex paths and relationships, for example, direct and indirect effects of the independent variable on the dependent variable and bidirectional relationships. This is useful because decoding proficiency could potentially contribute to L2 reading comprehension directly and also indirectly through enhancing vocabulary knowledge. Furthermore, the technique uses confirmatory factor analysis to assess the measurement model (i.e., whether the observed variables, say two

measures of decoding: Woodcock's test and TOWRE, touch on the same latent variable, say decoding ability) and can also account for any measurement error when analysing the relationships between the variables included in the model. Thus, these studies would not only validate findings from prior studies employing multiple regression techniques, but also provide more rigorous, and potentially more comprehensive, understanding of the contribution of L2 decoding proficiency to L2 reading comprehension performance.

7.1.5. A meta-analysis of the contribution of L2 decoding proficiency to L2 reading comprehension and the factors that may moderate this contribution.

The contribution of second language decoding proficiency to second language reading comprehension appears to vary somewhat considerably in the small but growing body of L2 literature. This variability makes it difficult to draw conclusions about the extent to which L2 decoding skill contributes to L2 reading comprehension. For this reason, it is essential to synthesize existing studies and explore the factors that may have moderated this contribution. Similar work has been undertaken in the L1 context (e.g., García and Cain, 2014) and which mainly focused on the moderating effects of assessment tools (for example, the decoding testing method) and learners' characteristics (for example, age, listening comprehension level, etc.). Similar meta-analytical studies are needed in the L2 context to rigorously examine whether this variability is indeed explained by moderating factors. These factors may include (a) the L1WS of the participants (e.g, congruent L1 and L2WSs vs. incongruent L1 and L2WSs); (b) the decoding and reading comprehension assessment tools used (for example, real words vs. pseudowords in the case of decoding, and timed vs. untimed passages in the case of L2 reading comprehension); (c) the proficiency level of the L2 learner (low proficiency vs. high proficiency); and most importantly, and often overlooked by previous L1 meta-analyses, (d) the analytical method adopted, and particularly, whether or not the contribution of decoding was assessed independently of that of other key predictors of reading comprehension.

7.1.6. Longitudinal studies into the contribution of L2 decoding proficiency to L2 reading comprehension performance

A natural progression of the present work is to carry out longitudinal studies to rigorously assess the impact of L2 decoding skill on L2 reading comprehension performance. Future longitudinal studies could adopt observational/correlational designs or, perhaps ideally, experimental designs. In the former, participants would be tested on the study measures (e.g., measures of L2 decoding and L2 reading comprehension) at two or more time points: the beginning of the study (time 1) and after a period of time (time 2). In the latter design, the researcher would randomly assign participants to treatment (decoding instruction) and control

(no decoding instruction) groups and assess the groups' decoding and reading comprehension skills pre and post intervention. When implemented adequately, longitudinal studies are more robust for establishing causal relations than cross-sectional ones. Thus, it would be valuable to see whether or not future longitudinal research confirms results found in the existing cross-sectional literature.

7.1.7. The contribution (or lack thereof) of L2 decoding proficiency to L2 listening comprehension performance

To the researcher's knowledge, no study has tested the impact of decoding proficiency on listening comprehension performance in first language or second language contexts. Absence of such investigations in the L1 context is reasonable, given that oral proficiency is normally developed prior to reading instruction. However, in the second language context, listening comprehension ability and decoding ability develop more or less concurrently (Yamashita and Shiotsu, 2015). In light of this, and reflecting on the researcher's own personal experience as a second language user of English, it can be argued that L2 decoding ability might predict variability in L2 listening comprehension performance. To illustrate, the researcher acquired numerous English words (i.e., word meanings and their respective phonological form) by exposure to written English texts. However, on occasions, her decoding – and hence stored pronunciation – of these new words was not accurate. When she was later exposed to the oral form of these words (for example, in a TV show, social interaction with English speakers, etc.), the researcher, being an advanced user of English, was often able to realise that her self-acquired pronunciation of the word was not accurate with minimal, if any, disruption to her listening comprehension. However, this might not be the case for beginner and possibly intermediate learners of English. Consider, for example, the case of a beginner learner who incorrectly decoded <choir> as /tʃɒɪr/, but who nonetheless has a correct understanding of the meaning of the word. When exposed to the correct phonological form of the word /kwaɪər/, this learner might not be able to access the previously acquired meaning of the word, potentially disrupting their listening comprehension. Therefore, it would be interesting to see studies that examine the contribution of L2 decoding proficiency to L2 listening comprehension performance.

References

Abdoh, E. M. A. (2011). *A study of the phonological structure and representation of first words in Arabic.* (PhD thesis). University of Leicester, Leicester, UK.

Abu-Rabia, S. (1996). The role of vowels and context in the reading of highly skilled native Arabic readers. *J Psycholinguist Res, 25*(6), 629-641.

Abu-Rabia, S. (1997). Reading in Arabic orthography: the effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing, 9*(1), 65-78.

Abu-Rabia, S. (1998). Reading Arabic texts: effects of text type, reader type and vowelization. *Reading and Writing, 10*(2), 105-119.

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing, 14*(1-2), 39-59.

Abu-Rabia, S. (2002). Reading in a root–based–morphology language: the case of Arabic. *Journal of Research in Reading, 25*(3), 299-309.

Abu-Rabia, S., & Taha, H. (2006). Reading in Arabic orthography: Characteristics, research findings, and assessment. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy*. Mahwah, N.J.: L. Erlbaum Associates.

Adams, M. J. (1990). Beginning to read. In. Cambridge, Ma: MIT Press.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing, 19*(9), 933-958.

Aiken, L. S., & West, S. G. (1991). *Multiple regression : testing and interpreting interactions*. Newbury Park, Calif.: Sage.

Alderson, J. C. (1984). Reading in a foreign language: a reading problem or a language problem? In J. C. Alderson & A. H. Urq'uhart (Eds.), *Reading in a Foreign Language* . London: Longman.

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

Alghamdi, H. (2015). *The PGST revisited: the role of context in Arabic reading*. Paper presented at the UK Orthography Group Meeting, Bristol, United Kingdom.

Alhazmi, K., Milton, J., & Johnston, S. (2019). Examining 'vowel blindness' among native Arabic speakers reading English words from the perspective of eye-tracking. *System, 80*, 235-245.

Alsubaie, M. (2014). *An Exploration of Reading Comprehension Challenges in Saudi Arabian University EFL Students.* (PhD thesis). The University of Exeter, Exeter, UK.

Alsulaimani, A. (1990). Reading problems in Arab learners of English. *Unpublished PhD thesis. London University*.

Ayari, S. (1996). Diglossia and illiteracy in the Arab world. *Language, Culture and Curriculum, 9*(3), 243-253.

Azen, R., & Budescu, D. (2003). The dominance analysis approach for comparing predictors in multiple regression. *8*(2), 129.

Baluch, B. (1993). Lexical Decisions in Persian: A Test of the Orthographic Depth Hypothesis. *International Journal of Psychology, 28*(1), 19-29.

Baluch, B., & Besner, D. (1991). Visual word recognition: Evidence for strategic control of lexical and nonlexical routines in oral reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 644.

Bauer, T. (1996). Arabic writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems*. Oxford: Oxford University Press.

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand journal of public health, 25*(5), 464-469.

Bernhardt, E. (2000). Second-language reading as a case study of reading scholarship in the 20th century. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research*. Mahwah, NJ: Lawrence Erlbaum.
Bernhardt, E. (2005). Progress and procrastination in second langauge reading. *Annual Review of Applied Linguistics, 25*, 133-150.
Bernhardt, E., & Kamil, M. (1995). Interpreting Relationships between L1 and L2 Reading: Consolidating the Linguistic Threshold and the Linguistic Interdependence Hypotheses. *Applied Linguistics, 16*(1), 15-34. doi:10.1093/applin/16.1.15
Besner, D., & Smith, M. (1992). Basic processes in reading: Is the orthographic depth hypothesis sinking? In F. Ram & K. Leonard (Eds.), *Orthography, phonology, morphology, and meaning* (Vol. Volume 94, pp. 227-248): North-Holland.
Bjaalid, I. K., Høien, T., & Lundberg, I. J. S. J. o. P. (1997). Dual-Route and Connectionist Models: A Step Towards a Combined Model. *38*(1), 73-82.
Braun, M. T., & Oswald, F. (2011). Exploratory regression analysis: A tool for selecting models and determining predictor importance. *Behaviour Research Methods, 43*(2), 331-339.
Bridgeman, B. (1987). Is the dual-route theory possible in phonetically regular languages? *Behavioral and Brain Sciences, 10*(02), 331-332.
Brown, T. L., & Haynes, M. (1985). Literacy background and reading development in a second language. *New Directions for Child and Adolescent Development, 1985*(27), 19-34.
Carr, T. H., & Levy, B. A. (1990). *Reading and its development : component skills approaches*. San Diego: Academic Press.
Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of Thoracic Disease, 7*(11), E537-E540.
Cavalcanti, M. (1984). Investigating FL readin Performance through verbal protocols. *The ESPecialist*(10).
Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*: John Wiley & Sons.
Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research, 29*(1), 1-24.
Cobb, T. (2019). Compleat Lexical Tutor. Retrieved from https://www.lextutor.ca
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement, 20*(1), 37-46.
Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed. ed.). London: London : Routledge.
Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*. Oxford: Blackwell publishing.
Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education, 4*(1), 5-17.
Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100*(4), 589.
Cook, V., & Bassetti, B. (2005). An introduction to researching second language writing systems. In V. Cook & B. Bassetti (Eds.), *Second Language Writing Systems*. Clevedon, Buffalo: Multilingual Matters.
Coulmas, F. (1999). *The Blackwell encyclopedia of writing systems*. Oxford: Blackwell Publishers.
Coussé, E., Gillis, S., Kloots, H., & Swerts, M. (2004). *The Influence of the Labeller's Regional Background on Phonetic Transcriptions: Implications for the Evaluation of Spoken Language Resources.* Paper presented at the LREC.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, Calif.: Sage.
Cucchiarini, C. (1993). *Phonetic transcription: a methodological and empirical study*.
Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research, 49*(2), 222-251.
Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89): Cambridge University Press.
Cunningham, A., Stanovich, K., & Wilson, M. (1990). Cognitive variation in adult college students differing in reading ability.
Cunningham, J. W. (1993). Whole-to-part reading diagnosis. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 9*(1), 31-49.
Davis, J. N., & Bistodeau, L. (1993). How do L1 and L2 reading differ? Evidence from think aloud protocols. *The modern language journal, 77*(4), 459-472.
Doff, A., & Robinson, N. (2013). *English Unlimited Starter B. Combo with DVD-ROMs (2)*: Cambridge University Press.
Dreyer, L., & Katz, L. (1992). An examination of "the simple view of reading.". *National Reading Conference Yearbook, 41*, 169-175.
Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*: American Guidance Service.
Durgunoğlu, A. Y. (2006). Learning to read in Turkish. *Developmental Science, 9*(5), 437-439.
Ebad, R. (2014). The Role and Impact of English as a Language and a Medium of Instruction in Saudi Higher Education Institutions: Students-Instructors Perspective. *Studies in English Language Teaching, 2*(2), 140.
Egan, W. J., & Morgan, S. L. (1998). Outlier detection in multivariate analytical chemical data. *Analytical chemistry, 70*(11), 2372-2379.
Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393-447.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second langauge: A Psychometric Study. *Studies in Second Language Acquisition, 27*(2), 141-172.
Ellis, R. (2008). *The study of second language acquisition* (2nd ed. ed.). Oxford: Oxford : Oxford University Press.
Emery, H. H. (2005). *An investigation into the nature and causes of reading and spelling errors made by Arab ESL learners.* Cardiff University, PhD Thesis.
Enders, C. K. (2003). Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychological methods, 8*(3), 322.
Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis : verbal reports as data* (Rev. ed.). Cambridge, Mass.: MIT Press.
Erler, L. (2003). *Reading in a foreign language : near-beginner adolescents' experiences of French in English secondary schools.* PhD thesis, University of Oxford.
Erler, L., & Macaro, E. (2011). Decoding ability in French as a foreign language and language learning motivation. *The modern language journal, 95*(4), 496-518.
Field, A. P. (2015). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.
Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review, 23*(4), 553-576.
Fox, J. (1991). *Regression diagnostics: An introduction* (Vol. 79): Sage.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German-and English-speaking children. *Scientific Studies of Reading, 2*(1), 31-54.
Frost, R. (2006). Becoming literate in Hebrew: the grain size hypothesis and Semitic orthographic systems. *Developmental Science, 9*(5), 439-440.
Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of experimental psychology. Human perception and performance, 13*(1), 104.
García, J. R., & Cain, K. (2014). Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English. *Review of Educational Research, 84*(1), 74-111.
Georgiou, G. K., Das, J. P., & Hayward, D. (2008). Revisiting the "Simple View of Reading" in a Group of Children With Poor Reading Comprehension. *Journal of Learning Disabilities, 42*(1), 76-84. doi:10.1177/0022219408326210
Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430.
Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6*(4), 126-135.
Goswami, U., Ziegler, J. C., Dalton, L., & Schnieder, W. (2003). Nonword Reading across Orthographies: How Flexible Is the Choice of Reading Units? *Applied Psycholinguistics, 24*(2), 235-247.
Gough, P. (1972). One second of reading. *Visible Language, 6*(4), 291-320.
Gough, P., & Hillinger, M. L. (1980). Learning to Read: an Unnatural Act. *Bulletin of the Orton Society, 30*, 179-196.
Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education, 7*(1), 6-10. doi:10.1177/074193258600700104
Grabe, W. (2009). *Reading in a second language : moving from theory to practice*. Cambridge: Cambridge University Press.
Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed. ed.). Harlow: Harlow : Longman.
Greenberg, D., Ehri, L. C., & Perin, D. (2002). Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading, 6*(3), 221-243.
Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology, 32*(3), 282-313.
Hadi, A. S., & Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association, 88*(424), 1264-1272.
Hamada, M., & Koda, K. (2008). Influence of First Language Orthographic Experience on Second Language Decoding and Word Learning. *Language Learning, 58*(1), 1-31.
Hamada, M., & Koda, K. (2010). The Role of Phonological Decoding in Second Language Word-Meaning Inference. *Applied Linguistics, 31*(4), 513-531.
Hanson, V. L., Goodell, E. W., & Perfetti, C. (1991). Tongue-twister effects in the silent reading of hearing and deaf college students. *Journal of Memory and Language, 30*(3), 319-330.
Hayes, A. F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach. New York, NY: The Guliford Press.
Hayes-Harb, R. (2006). Native Speakers of Arabic and ESL Texts: Evidence for the Transfer of Written Word Identification Processes. *TESOL quarterly, 40*(2), 321-339.

Haynes, M., & Carr, T. (1990). Writing system background and second language reading: A component skills analysis of English reading by native speaker-readers of Chinese. In T. H. Carr & B. A. Levy (Eds.), *Reading and its development : component skills approaches*. San Diego: Academic Press.
Holm, A., & Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition, 59*(2), 119-147.
Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160.
IBM Corp. (2015). *IBM SPSS Statistics for Macintosh, Version 23.0.* . Armonk, NY: IBM Corp.
Jeon, E. H. (2011). Contribution of Morphological Awareness to Second-Language Reading Comprehension. *The modern language journal, 95*(2), 217-235.
Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212.
Jiang, X. (2016). The role of oral reading fluency in ESL reading comprehension among learners of different first language backgrounds. *Reading Matrix: An International Online Journal, 16*(2), 227-242.
Jiang, X., Sawaki, Y., & Sabatini, J. (2012). Word Reading Efficiency, Text Reading Fluency, and Reading Comprehension Among Chinese Learners of English. *Reading Psychology, 33*(4), 323-349.
Johnson, J. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Journal of Multivariate behavioral research, 35*(1), 1-19.
Johnson, J. W., & LeBreton, J. M. J. O. r. m. (2004). History and use of relative importance indices in organizational research. *7*(3), 238-257.
Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing, 19*(4), 339-361.
Joshi, R. M., & Aaron, P. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85-97.
Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In F. L. RAM, K. (Ed.), *Orthography, Phonology, morphology, and meaning* (Vol. 94, pp. 67-84): North Holland.
Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10*(4), 363-380.
Kieffer, M. J., & Lesaux, N. K. (2012). Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers. *Language Learning, 62*(4), 1170-1204.
Kim, H.-Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics, 38*(1), 52-54.
Kingston, A. J. (1961). A conceptual model of reading comprehension. In J. Emery P. Bliesmer and Albert J. Kingston (Ed.), *Phases of college and other adult reading programs. Tenth Yearbook of the National Reading Conference.*
Kirby, J. R., & Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy, 42*(2), 75-82.
Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London: Routledge.
Koda, K. (1998). The role of phonemic awareness in second language reading. *Second Language Research, 14*(2), 194-215.
Koda, K. (2005). *Insights into second language reading: a cross-linguistic approach*. Cambridge: Cambridge University Press.

Koda, K. (2008). Impacts of prior literacy experience on second-language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: cross-linguistic relationships in first- and second-language literacy development*. London: Routledge.
Koda, K. (2013). Learning to Read in New Writing Systems. In *The Encyclopedia of Applied Linguistics*: Blackwell Publishing Ltd.
Kraha, A., Turner, H., Nimon, K., Zientek, L., & Henson, R. J. F. i. p. (2012). Tools to support interpreting multiple regression in the face of multicollinearity. *3*, 44.
Kremmel, B., Brunfaut, T., & Alderson, J. C. (2015). Exploring the Role of Phraseological Knowledge in Foreign Language Reading. *Applied Linguistics*, amv070.
LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323.
Lebreton, J. M., Ployhart, R. E., & Ladd, R. T. J. O. R. M. (2004). A Monte Carlo comparison of relative importance methodologies. *7*(3), 258-282.
LeBreton, J. M., & Tonidandel, S. J. J. o. A. P. (2008). Multivariate relative importance: Extending relative weight analysis to multivariate criterion spaces. *93*(2), 329.
Lee, J. W., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL quarterly, 31*(4), 713-739.
Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196-228.
Leys, C., Klein, O., Dominicy, Y., & Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology, 74*, 150-156.
Li, S. (2019). The effects of phonics instruction on L2 phonological decoding and vocabulary learning. In R. Woore & C. Walter (Eds.).
Logan, G. D. (1997). Automaticity and reading: perspectives from the instance theory of automatization. *Reading & Writing Quarterly, 13*(2), 123-146.
Lorenzo-Seva, U., Ferrando, P. J., & Chico, E. J. B. r. m. (2010). Two SPSS programs for interpreting multiple regression results. *42*(1), 29-35.
Martínez, A. C. L. (2009). A state-of-the-art review of background knowledge as one of the major factors that influence reading comprehension performance. *Elia: Estudios de lingüística inglesa aplicada*(9), 31-57.
McNeil, L. (2012). Extending the compensatory model of second language reading. *System, 40*(1), 64-76.
Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjær, & J. Williams (Eds.), *Performance and competence in second language acquisition*. Cambridge: Cambridge : Cambridge University Press.
Mellard, D., Anthony, J., & Woods, K. (2012). Understanding oral reading fluency among adults with low literacy: dominance analysis of contributing component skills. *An Interdisciplinary Journal, 25*(6), 1345-1364.
Ministry of Education. (2017). Breakdown of foundation year students across higher education institutions. Retrieved from http://departments.moe.gov.sa/PlanningInformation/RelatedDepartments/Educationstatisticscenter/EducationDetailedReports/Docs/table1-2.html
Nassaji, H. (2003). Higher–level and lower–level text processing skills in advanced ESL reading comprehension. *The modern language journal, 87*(2), 261-276.
Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching, 47*(01), 1-37.

Nassaji, H., & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics, 20*(02), 241-267.
Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47-73. doi:10.1080/19404158.2019.1609272
Oller, D. K., & Ramsdell, H. L. (2006). A weighted reliability measure for phonetic transcription. *Journal of Speech, Language, and Hearing Research, 49*(6), 1391-1411.
Olson, G. M., Duffy, S. A., & Mack, R. L. (2018). Thinking-Out-Loud as a Method for Studying 11 Real-Time Comprehension Processes. *New methods in reading comprehension research*, 253.
Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of Reference for the Assessment of Learning Disabilities: New Views on Measurement Issues*. Baltimore: Paul H. Brookes Publishing Co.
Paris, S. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184-202.
Pasquarella, A., Gottardo, A., & Grant, A. (2012). Comparing Factors Related to Reading Comprehension in Adolescents Who Speak English as a First (L1) or Second (L2) Language. *Scientific Studies of Reading, 16*(6), 475-503.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research : explanation and prediction* (3rd ed.). London: Harcourt Brace College Publishers.
Perfetti, C. (1985). *Reading ability*: Oxford University Press.
Perfetti, C. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*(1), 3-24.
Perfetti, C. (2010). Decoding, vocabulary, and comprehension. In M. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 291-303). New York: The Guilford Press.
Perfetti, C., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*(1), 59-70.
Perfetti, C., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L.-H. (2007). Reading in Two Writing Systems: Accommodation and Assimilation of the Brain's Reading Network. *Bilingualism: Language and Cognition, 10*(2), 131-146.
Perfetti, C., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "Universal" Phonological Principle. In F. Ram & K. Leonard (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227-248): North-Holland.
Pollatsek, A. (2015). The role of sound in silent reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading.* (pp. 185-201). Oxford.
Porpodas, C., Pantelis, S., & Hantziou, E. (1990). Phonological and lexical encoding processes in beginning readers: Effects of age and word characteristics. *Reading and Writing, 2*(3), 197-208.
Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language testing, 21*(1), 28-52.
R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria.
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, N.J.: Prentice Hall.
Rea, D. (2010). *English unlimited. A2 elementary*. Cambridge: Cambridge University Press.

Ryan, A., & Meara, P. (1991). The case of the invesible vowels: Arabic speakers reading English words. *Reading in a foreign language, 7*(2), 531-540.
Saiegh-Haddad, E., & Henkin-Roitfarb, R. (2014). The structure of Arabic language and orthography. In E. Saiegh-Haddad & R. M. Joshi (Eds.), *Handbook of Arabic literacy: insights and perspectives* Berlin: Springer.
Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System, 40*(1), 24-36.
Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign Language Reading Anxiety. *The modern language journal, 83*(2), 202-218.
Savage, R., & Wolforth, J. (2007). An Additive Simple View of Reading Describesthe Performance of Good and Poor Readersin Higher Education. *Exceptionality Education International, 17*(2), 243-268.
Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research, 8*(1), 3-15.
Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing, 18*(1), 55-88.
Seidenberg, M. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, Phonology, morphology, and meaning* (Vol. 94, pp. 85-118). Burlington Elsevier.
Seidenberg, M. S. (2007). Connectionist models of reading. In M. G. Gaskell & G. T. M. Altmann (Eds.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press.
Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review, 96*(4), 523-568.
Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.
Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In *Advances in child development and behavior* (Vol. 36, pp. 31-82): Elsevier.
Share, D. L. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127.
Shiotsu, T. (2010). *Components of L2 reading: Linguistic and processing factors in the reading test performances of Japanese EFL learners* (Vol. 32). Cambridge: Cambridge University Press.
Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language testing, 24*(1), 99-128.
Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics, 5*(3), 225-279.
Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency Has a Role in the Simple View of Reading. *Scientific Studies of Reading, 17*(2), 108-133.
Smith, F. (1978). *Understanding reading : a psycholinguistic analysis of reading and learning to read* (2nd ed.). New York: Holt, Rinehart and Winston.
Sorg, B. A., & Whitney, P. (1992). The effect of trait anxiety and situational stress on working memory capacity. *Journal of Research in Personality, 26*(3), 235-241.
Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology, 27*(2), 217-235.

Stanovich, K. (1980). Toward an Interactive-Compensatory Model of individual differences in the development of reading Fluency. *Reading Research Quarterly, 16*(1), 32-71.
Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360-407.
Stanovich, K. (2000). *Progress in understanding reading : scientific foundations and new frontiers*. London: Guilford Press.
Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, N.J.
London: Erlbaum.
Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Boston
London: Pearson.
Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition, 20*(3), 303-313.
Tan, A., Moore, D. W., Dixon, R. S., & Nicholson, T. (1994). Effects of training in rapid decoding on the reading comprehension of adult ESL learners. *Journal of Behavioral Education, 4*(2), 177-189.
Tannenbaum, R. J., & Wylie, E. C. (2008). linking English language test scores onto the Common European Framework of Reference: an application of standard-setting methodology. *ETS Research Report Series, 2008*(1), i-75.
Taouka, M., & Coltheart, M. (2004). The cognitive processes involved in learning to read in Arabic. *Reading and Writing, 17*(1-2), 27-57.
Tighe, E. L., & Schatschneider, C. (2014). A dominance analysis approach to determining predictor importance in third, seventh, and tenth grade reading comprehension skills. *Reading and Writing, 27*(1), 101-127.
Tilbury, A., Clementson, T., Hendra, L., & Rea, D. (2016). English Unlimited B1. Pre-intermediate: Coursebook with e-Portfolio DVD-ROM. Cambridge University Press.
Tonidandel, S., & LeBreton, J. (2011). Relative Importance Analysis: A Useful Supplement to Regression Analysis. *Journal of Business and Psychology, 26*(1), 1-9.
Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business Psychology, 30*(2), 207-216.
Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE [multimedia] : test of word reading efficiency*. Austin, Tex.: Pro-Ed.
Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE 2 : test of word reading efficiency* (Second ed.). Austin, Texas: Pro-Ed.
Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of experimental psychology, 124*(2), 107.
Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), *Orthographies and reading : perspectives from cognitive psychology, neuropsychology, and linguistics*. Hillsdale, N.J: Lawrence Erlbaum Associates.
Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181-198.
Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and writing, 25*(8), 1805-1818. doi:10.1007/s11145-011-9346-3
Walter, C. (2007). First- to second-language reading comprehension: not transfer, but access1. *International Journal of Applied Linguistics, 17*(1), 14-37.

Walter, C. (2008). Phonology in second language reading: Not an optional extra. *TESOL quarterly, 42*(3), 455-474.
Wang, M., & Koda, K. (2007). Commonalities and Differences in Word Identification Skills Among Learners of English as a Second Language. *Language Learning, 57*, 201-222.
Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: a comparison of Korean and Chinese English L2 learners. *Cognition, 87*(2), 129-149.
Wiederholt, J. L., & Blalock, G. (2000). *Gray Silent Reading Tests GSRT*: Pro-Ed.
Wimmer, H. (2006). Don't neglect reading fluency! *Developmental Science, 9*(5), 447.
Woodcock, R. W. (1987). *Woodcock reading mastery tests*. Circle Pines, MN: American Guidance Service
Woore, R. (2010). Thinking aloud about L2 decoding: an exploration of the strategies used by beginner learners when pronouncing unfamiliar French words. *Language Learning Journal, 38*(1), 3-17.
Woore, R. (2011). *Investigating and developing beginner learners' decoding proficiency in second language French: an evaluation of two programmes of instruction.* Oxford University, UK,
Woore, R. (2013). Beginner learners' progress in decoding L2 French: Transfer effects in typologically similar L1-L2 writing systems. *Writing Systems Research*(ahead-of-print), 1-23.
Woore, R. (2016). Learners' pronunciations of familiar and unfamiliar French words: what can they tell us about phonological decoding in an L2? *The Language Learning Journal*, 1-14.
Woore, R., Graham, S., Porter, A., Courtney, L., & Savory, C. (2018). *Foreign Language Education: Unlocking Reading (FLEUR) - A study into the teaching of reading to beginner learners of French in secondary school*.
Yamashita, J., & Shiotsu, T. (2015). Comprehension and Knowledge Components That Predict L2 Reading: A Latent-Trait Approach. *Applied Linguistics*. doi:10.1093/applin/amu079
Zhang, D. (2012). Vocabulary and grammatical knowledge in L2 reading comprehension: A structural equation modeling study. *The modern language journal, 96*, 554-571.
Zhang, D., & Koda, K. (2012). Contribution of Morphological Awareness and Lexical Inferencing Ability to L2 Vocabulary Knowledge and Reading Comprehension among Advanced EFL Learners: Testing Direct and Indirect Effects. *Reading and Writing: An Interdisciplinary Journal, 25*(5), 1195-1216.
Zhao, A. (2013). Foreign language reading anxiety: Chinese as a foreign language in the United States. *The modern language journal, 97*(3), 764.
Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin, 131*(1).
Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical Words are Read Differently in Different Languages. *Psychological Science, 12*(5), 379-384.